Self-licensing

Blanken, Irene

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Self-Licensing: When and Why People Give in to Temptation

Irene Blanken

Self-Licensing: When and Why People Give in to Temptation

PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan Tilburg University

op gezag van de rector magnificus,

prof. dr. E. H. L. Aarts,

in het openbaar te verdedigen

ten overstaan van een door het college

voor promoties aangewezen commissie

in de aula van de Universiteit

op vrijdag 27 november 2015 om 14.15 uur

door Irene Blanken,

geboren op 17 juni 1987 te Ouderkerk

Promotor	prof. dr. Marcel Zeelenberg
Copromotor	dr. ing. Niels van de Ven

Promotiecommissie	dr. Barbara M. E. Briers
	prof. dr. Wilhelm Hofmann
	dr. Tila M. Pronk
	dr. Marijke van Putten
	prof. dr. Denise T. D. de Ridder

Table of contents

CHAPTER 1

Introduction

In the summer of 2014, the world was captivated by the Ice Bucket Challenge. People nominated each other via social media to post videos in which they dump a bucket filled with cold water and ice cubes on their head to promote awareness of amyotrophic lateral sclerosis (ALS) disease and encourage donations to ALS-related charities. As a result of this viral trend, the national ALS Association raised over $94 million in July and August 2014. At first glance, this social media hype seems a brilliant way of raising money for charity. However, many people worried that the worldwide donations to ALS charities would be at the cost of other charities with greater healthcare value, because they assumed that people who had just donated money to ALS would be less likely to subsequently donate money to other charities.

The criticism on the Ice Bucket Challenge closely reflects the phenomenon that is investigated in the current dissertation, which is called *self-licensing*. Self-licensing implies that people allow themselves to engage in less desirable behavior *because* they previously performed good behavior. For people who participated in the Ice Bucket Challenge, this infers that donating money to an ALS charity can result in fewer charitable actions in the direct future. This dissertation reports studies that examine the self-licensing effect, which are presented in Chapter 2 to 7. Before turning to these studies, an overview is provided on what self-licensing is, why it is important to study self-licensing, and what is missing in the current literature on self-licensing.

What is self-licensing?

Self-licensing theory posits that people who behave in a good way, later feel that they are permitted to engage in undesirable behavior (Miller & Effron, 2010). Thus, a prior good deed can provide a license that allows one to perform undesirable behavior later on. The term *self-licensing* was introduced by Monin and Miller (2001), who proposed that people who first establish *nonracist credentials* through acting unprejudiced are subsequently licensed to display prejudiced behavior. Since then, more than 50 separate studies on self-licensing have been published. Self-licensing is interesting and relevant. It is interesting because it can help explain why people engage in "bad" behavior and because it makes predictions about people's internal accounting systems of good and bad behaviors. Self-licensing is relevant because it can both help and hurt

people. It can help people to loosen up a little bit, but it can hurt individual and social goals.

Self-licensing related to individual behaviors

Research suggests that behaving in a good way can result in negative behavior that has harmful consequences at the individual level. On a daily base, individuals are confronted with numerous goal-incongruent temptations, such as unhealthy high caloric foods when one is dieting, spending money on luxurious items while trying to save money for retirement, and undertaking fun activities instead of working. Self-licensing can lead to indulgence in these various domains. Khan and Dhar (2006), who conducted the first studies on self-licensing related to individual behaviors, suggested that "a prior choice, which activates and boosts a positive self-concept, subsequently licenses the choice of a more self-indulgent option" (p. 260). Studies on self-licensing related to individual behaviors found that prior goal progress on weight loss objectives (Fishbach & Dhar, 2005), prior restraint (Mukhopadhyay, Sengupta, & Ramanathan, 2008), and perceived prior effort (De Witt-Huberts, Evers, & De Ridder, 2012) can result in more indulgent food choices. In addition, intended virtuous behaviors (Khan & Dhar, 2006) and prior shopping restraint (Mukhopadhyay & Johar, 2009) can lead to preferences for self-indulgent hedonic over functional products. Furthermore, perceived goal progress on academic course work can result in greater interest in pursuing non-academic activities (Fishbach & Dhar, 2005). Thus, self-licensing is related to behaviors that are incongruent with personal goals such as dieting, saving money, and working.

Self-licensing related to social behaviors

Research suggests that behaving in a good way can also result in negative behavior that has harmful consequences at the societal level. For instance, participants who recalled prior moral behavior subsequently displayed less prosocial intentions towards others and cheated more on a computer task compared to participants who recalled prior immoral behavior (Jordan, Mullen, & Murnighan, 2011). In addition, participants who established their unprejudiced attitudes through endorsing

president Obama (Effron, Cameron, & Monin, 2009) or through preferring a black over a white person for a consulting firm job (Monin & Miller, 2001) were subsequently more likely to make prejudiced decisions than participants who did not establish their unprejudiced attitudes. Furthermore, participants who bought more ecofriendly than regular products in a virtual shopping paradigm subsequently offered less money to other participants and stole more money from the experimenter compared to participants who bought more regular products (Mazar & Zhong, 2010). Self-licensing related to social behaviors is often referred to as *moral licensing* (for an extensive overview on moral licensing, see Chapter 2), which Merritt, Effron, and Monin (2010, p. 344) define as:

> When people are under the threat that their next action might be (or appear to be) morally dubious, individuals can derive confidence from their past moral behavior, such that an impeccable track record increases their propensity to engage in otherwise suspect actions.

It has been suggested that moral licensing can be interpreted as part of a larger moral self-regulation framework, where internal balancing of moral self-worth and the costs associated with prosocial behavior determine whether one will display (im)moral behavior (e.g., Sachdeva, Iliev, & Medin, 2009). According to this framework, establishing a moral image of oneself allows a person to display an immoral action without the fear of losing that moral image (leading to licensing effects). On the contrary, establishing an immoral image of oneself through appearing immoral to others creates a need for subsequent positive actions to restore the moral image (leading to compensation or cleansing effects; Zhong & Liljenquist, 2006).

Taken together, self-licensing occurs in various domains, both with respect to the origin of the license and with the behavior that it instigates. Self-licensing can play an important role in individual self-regulatory domains. When people feel good about their prior actions, they become more likely to perform goal-incongruent activities, such as consuming unhealthy foods, spending money on luxurious items, or engaging in fun activities rather than studying. The negative behavior self-licensing results in can also have harmful consequences at the societal level, such as displaying prejudiced attitudes, acting dishonestly, and engaging in less virtuous behaviors.

Why is it important to study self-licensing?

Interestingly, self-licensing theory seems to conflict with a number of well-established psychological findings and theories that highlight humans' desire for cognitive consistency in their thoughts, feelings, and behavior (Abelson et al., 1986; Gawronski & Strack, 2012). For instance, balance theory (Heider, 1946), cognitive dissonance theory (Festinger, 1957), the foot-in-the-door effect (Freedman & Fraser, 1966), and the sunk cost effect (Arkes & Blumer, 1985) all highlight consistency as a main motivator in human behavior. In addition, self-perception theory asserts that the attitudes that people infer from observation of their prior behavior ultimately affect their subsequent actions (Bem, 1972). Bem would argue that people who perform a good deed would see themselves as being good and moral persons, which would therefore lead to future moral behavior as well. After all, people generally intend to maintain positive self-images and want to act and feel like good persons (Aronson, Cohen, & Nail, 1999; Schlenker, 1980; Steele, 1988). Thus, self-licensing seems to be at odds with a number of well-established findings and theories that predict that individuals aim to be (and like to appear) good and consistent in their behaviors, and especially so when past good behaviors have just been highlighted. It is vital to investigate why people in self-licensing situations do not seem to have the need for consistency, as this can explain when and why people deviate from their typical behavioral patterns.

Besides the theoretical contradiction between self-licensing and a number of established psychological theories, studying licensing also has relevant practical implications. As outlined in the previous section, licensing can lead to a broad spectrum of undesirable behaviors, both at the individual and societal level. Research on this topic thus gives important insights into people's motivation and behavior, and can ultimately prevent people from displaying goal-incongruent and vicious behavior.

What is currently unknown about self-licensing

Although a substantial number of studies on licensing has been published, much is still unknown about self-licensing. To date, it is unclear what the size of the licensing effect is, how robust the effect is, and what the important moderators of the effect are.

After all, moderators have to exist if both consistency and licensing effects occur. In addition, it is not clear whether the effects obtained in the lab are easy to replicate, since a large diversity of paradigms has been used to study licensing. There is not one solid paradigm that is often used to test self-licensing and its potential moderators in experimental settings.

Moreover, only a very limited amount of research on self-licensing has been conducted outside of laboratory settings. Therefore, it is unknown if, when, and how self-licensing operates in daily life. Furthermore, self-licensing shows some close resemblances with theorizing on motivated reasoning (Kunda, 1990), which posits that when individuals have a preferred conclusion, they attempt to construct justifications that are supportive of that particular conclusion. The basic licensing idea that bad behavior can follow from good behavior may also exist if people are tempted to engage in the bad behavior and use their prior good behavior to justify engaging in it. It is unclear whether motivated reasoning plays a role in the process of self-licensing, because it has not explicitly been tested whether self-licensing can be initiated by the temptation to display undesirable behaviors (see Chapter 5 and 6 for further elaboration on this). Finally, little is known about the motivational consequences of having (licensing-related) reasons for displaying goal-incongruent behavior. The present dissertation aims to fulfill these gaps in the current literature on self-licensing.

Overview of the current dissertation

The dissertation consists of two parts. The first part of this presents a state-of-the-art overview of self-licensing. In this part, the effect size of licensing, the robustness and replicability of the effect, and theoretical moderators are investigated. The second part presents novel perspectives on self-licensing. This part investigates in which ways self-licensing can be triggered in daily life, an alternative account of self-licensing, and the consequences of having (licensing-related) reasons for displaying goal-incongruent behavior.

All empirical chapters are written as separate papers and could therefore be read independently and in any order. As a result, however, there may be some overlap

between the different chapters. Since the empirical chapters are written together with my co-authors, the word 'we' is used throughout these chapters.

Part I: The State-of-the-Art of Self-Licensing

Chapter 2. A Meta-Analytic Review of Moral Licensing

This chapter provides a state-of-the art overview of moral licensing. A meta-analysis including 91 studies that compare a licensing condition with a control condition (7397 participants) is conducted. The magnitude of the moral licensing effect is estimated and potential moderators of the effect are investigated. Based on this analysis, practical and theoretical recommendations for scholars studying the moral licensing effect are provided.

Chapter 3. A P-Curve Analysis of Moral Licensing

Chapter 3 tests the evidential value of the moral licensing effect through conducting a *p*-curve analysis (Simonsohn, Nelson, & Simmons, 2014a,b) on the published moral licensing tests that were included in the meta-analysis from Chapter 2. The *p*-curve analysis provides a conclusive answer on whether the established moral licensing effect is true, or the studies merely reflect selective reporting. Unlike the meta-analysis, the *p*-curve analysis is able to provide an effect size estimate that is not influenced by publication bias. Implications for research on moral licensing are discussed.

Chapter 4. Three Attempts to Replicate the Moral Licensing Effect

Chapter 4 focuses on the replicability of the licensing effect. This chapter includes three studies that attempt to replicate the moral licensing effect that was previously established by Sachdeva et al. (2009). In their important contribution to the licensing literature, they found that writing about positive traits led to lower donations to charity and decreased cooperative behavior. They also found the opposite effect (i.e., moral cleansing): Writing about negative traits led to more donations to charity and increased cooperative behavior. Study 4.1 and 4.2 aim to replicate their findings in student samples (95% power based on the original effect), and Study 4.3 aims to

replicate the effect in a more generalizable Amazon Mechanical Turk sample (95% power based on the effect size from the meta-analysis in Chapter 2).

Part II: Novel Perspectives on Self-Licensing

Chapter 5. Two Ways in Which Self-Licensing can be Triggered

Chapter 5 tests the proposition that there are two different ways in which self-licensing can be triggered; 1) a good deed can permit one to display undesirable behavior (*good deed self-licensing*) and 2) the temptation to display undesirable behavior can initiate a search for a license (*temptation-based self-licensing*). This is tested through observations of the natural occurrence of self-licensing. Study 5.1 explores the self-licensing effects people recall from their daily lives and investigates to what extent these effects reflect good deed self-licensing versus temptation-based self-licensing. Study 5.2 manipulates these different ways in which self-licensing can be triggered and explores the different antecedents and consequences.

Chapter 6. Temptation-Based Reasoning

In Chapter 6, a temptation-based reasoning model is proposed and tested. This model posits that people interpret reasons for indulgence in a different light depending on how tempting behavior is. Specifically, Chapter 6 tests whether reasons for indulgence are seen as more acceptable when behavior is more tempting. Study 6.1 and 6.2 test whether people evaluate given reasons as more acceptable reasons to indulge when they are confronted with more tempting behavior. Study 6.3 and 6.4 test whether people interpret their own recalled prior good behaviors and frustrations as better reasons to indulge when behavior is more tempting. Based on the findings, an alternative account of self-licensing is proposed.

Chapter 7. Reasonable Reasons for Indulgence

Whereas Chapter 6 emphasizes how reasoning processes can facilitate indulgence, Chapter 7 focuses on the consequences of having different reasons for giving in to temptation (i.e., breaking a diet by unhealthy snacking). Specifically, the affective and motivational consequences of having different reasons for displaying undesirable

behavior (i.e., breaking a diet) are investigated. This chapter builds on previous literature that suggests that health-interventions should focus on the most common reasons people have for breaking their diet. The focus from the quantitative aspects of breaking a diet (which reasons are most common?) is shifted to the qualitative aspects (what are the consequences of breaking a diet for various reasons?). Study 7.1 tests whether reasonable reasons for unhealthy snacking have less negative consequences for further goal striving than unreasonable reasons for unhealthy snacking. Study 7.2a and 7.2b test whether reasons for unhealthy snacking that have been identified in the literature differ in how reasonable they are. Finally, Study 7.3 investigates whether the two categories of reasons for unhealthy snacking that are most common (enjoying a special occasion versus opportunity-induced eating) have different motivational consequences. This chapter serves as a first step in exploring the consequences of having (licensing-related) reasons for displaying goal-incongruent behavior.

Chapter 8. Discussion

Chapter 8 integrates and discusses the findings from the empirical chapters. This chapter provides a summary of the findings throughout this dissertation and discusses the theoretical implications of these findings. Furthermore, the findings are placed in a broader theoretical framework. Finally, practical implications and future directions are discussed.

Part I: The State of The Art of Self-Licensing

CHAPTER 2

A Meta-Analytic Review of Moral Licensing

Moral licensing refers to the effect that when people initially behave in a moral way, they are later more likely to display behaviors that are immoral, unethical, or otherwise problematic. In this chapter we provide a state-of-the-art overview of moral licensing by conducting a meta-analysis of 91 studies (7397 participants) that compare a licensing condition to a control condition. Based on this analysis the magnitude of the moral licensing effect is estimated to be a Cohen's *d* of 0.31. We tested potential moderators and found that published studies tend to have larger moral licensing effects than unpublished studies. We found no empirical evidence for other moderators that were theorized to be of importance. The effect size estimate implies that studies require many more participants to draw solid conclusions about moral licensing and its possible moderators.

This chapter is based on Blanken, I., Van de Ven, N., & Zeelenberg, M. (2015a). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, *41*, 540-558. doi: 10.1177/0146167215572134

Moral licensing theory posits that people who initially behave in a moral way can later display behaviors that are immoral, unethical, or otherwise problematic (e.g., Merritt, Effron, & Monin, 2010). For example, someone who has just spent some time volunteering for the local community center might later find it more acceptable to "forget" to report some additional income when filling out the tax return. The idea of moral licensing theory is that the prior good deed provides a 'license', that allows one to perform morally questionable behavior later on. To date, various studies on moral licensing have been performed, a substantial subset of which has been published. However, the magnitude of the effect and the specific conditions under which moral licensing is likely to occur remain unclear. Therefore, in this chapter, we provide a comprehensive overview by performing a meta-analysis across all the available studies on moral licensing.

The mere existence of moral licensing, in which one allows oneself to engage in less moral behavior after a prior good deed, seems to be at odds with a number of well-established psychological findings and theories that stress consistency in behavior. Self-perception theory, for example, asserts that people infer their attitudes from observations of their own behavior which ultimately affects their subsequent behavior (Bem, 1972). People who perform a good deed would thus see themselves as being a good and moral person, which would therefore lead to future moral behavior as well. In addition, balance theory (Heider, 1946), cognitive dissonance theory (Festinger, 1957), the foot-in-the-door effect (Freedman & Fraser, 1966), and the sunk cost effect (Arkes & Blumer, 1985), all highlight consistency as an essential motivator of human behavior (for reviews see Abelson et al., 1968; Gawronski & Strack, 2012). Important social psychological theories thus predict that people like to be (and like to appear) good and moral in their actions, and especially so when past moral behavior has just been highlighted, as that makes it important to appear consistent. However, the moral licensing effect has been reported in many domains, including job hiring (Cascio & Plant, 2015; Monin & Miller, 2001), ambiguous racist attitudes (Choi, Crandall, & La, 2014; Effron, Cameron, & Monin, 2009; Effron, Monin, & Miller, 2012; Mann, & Kawakami, 2012), donations to charity (Conway & Peetz, 2012; Sachdeva, Iliev, & Medin, 2009) consumer behavior (Khan & Dhar, 2006), and dishonest behavior

(Jordan, Mullen, & Murnighan, 2011; Mazar & Zhong, 2010). As moral licensing has these adverse consequences for such a wide range of behaviors, research on this topic can give important insights into people's motivation and behavior.

It has been suggested that moral licensing can be interpreted as part of a larger moral self-regulation framework. The idea is that internal balancing of moral self-worth and the costs associated with prosocial behavior determine whether one will display (im)moral behavior (e.g., Sachdeva et al., 2009). When the moral image of oneself is established, an immoral action is allowed without the fear of losing that moral image (leading to licensing). Conversely, when one appears immoral to others, subsequent positive actions are needed to restore the moral image (leading to compensation or cleansing). However, further research is needed before it can be concluded that a general 'balancing' mechanism is responsible for both the licensing and the cleansing effect (cf. Chapter 4). The focus of the current meta-analysis will be entirely on the moral licensing effect. A meta-analysis on moral licensing will help in painting a clearer picture on what licensing is and when it occurs, and therefore also forms a solid basis for exploring how the processes of moral licensing and moral cleansing relate.

Definitions of moral licensing

On a theoretical level, the process of moral licensing is defined as

> When people are under the threat that their next action might be (or appear to be) morally dubious, individuals can derive confidence from their past moral behavior, such that an impeccable track record increases their propensity to engage in otherwise suspect actions (Merritt et al., 2010, p. 344).

For our meta-analysis, we build upon the more operational definition or moral licensing that Merritt et al. (2010, p. 344) provide: "Past good deeds can liberate individuals to engage in behaviors that are immoral, unethical, or otherwise problematic, behaviors that they would otherwise avoid for fear of feeling or appearing immoral". Moral licensing can be regarded as an example of the broader category of psychological licensing which is "the perception that one's behavioral history, social context, or category membership permit one to legitimately do or say something that otherwise would discredit the self" (Miller & Effron, 2010, p. 116). For

example, being a member of a minority group can license one to criticize that group. Thus, moral licensing makes an appeal to someone's past good deeds, whereas the broader category of psychological licensing does not necessarily involve performing good behavior or displaying good intentions. Studies including this particular type of psychological licensing are not included in the meta-analysis, since we are interested in the behavioral consequences of acting morally.

Typical examples of moral licensing

Moral licensing occurs for both individual and social behaviors. It touches upon relevant everyday behaviors related to welfare, job hiring, ambiguous racial attitudes, charity donations, consumer purchases and green consumption. For instance, participants who established their non-prejudiced attitudes by endorsing president Obama (Effron et al., 2009) or through selecting a black person for a consulting firm job (Monin & Miller, 2001) were subsequently more likely to make pro-white judgments. In a similar vein, participants who previously received feedback that they were close to their goal of being regarded egalitarian towards black persons seated themselves farther away from a black confederate than participants who received feedback that they were not progressing toward their goal (Mann & Kawakami, 2012). Furthermore, participants who recalled their own moral actions subsequently displayed fewer prosocial intentions (Conway & Peetz, 2012; Jordan et al., 2011) and cheated more to get a higher payoff (Clot, Grolleau, & Ibanez, 2014; Jordan, et al., 2011) than participants who did not recall their moral actions. Moral licensing also plays a role in the area of interpersonal decision making: Previous ethical behavior licensed participants to offer less money to other participants (Cornelissen, Bashshur, Rode, & Le Menestrel, 2013a). In addition, participants who disclosed their conflict of interest to the other party subsequently gave more corrupt advice (Cain, Loewenstein, & Moore, 2005; 2010).

Moral licensing does not seem to be a within-domain phenomenon; the licensed behavior can also take place in a different area. For instance, participants who imagined that they would volunteer to spend time doing community service subsequently preferred hedonic over utilitarian products (Khan & Dhar, 2006).

Furthermore, in a virtual shopping paradigm, participants who bought ecofriendly products subsequently offered less money to another person in an ultimatum game and stole more money compared to participants who bought regular products (Mazar & Zhong, 2010).

The current meta-analysis

The first moral licensing study was published in 2001 (Monin & Miller, 2001). Since then, over 50 separate studies have been published in more than 20 articles. One reason for conducting the current meta-analysis was our observation that the moral licensing effect seems to conflict with one of the most established psychological findings that people want to be and appear consistent in their behavior. A second reason is based on our own research experiences and informal contacts with colleagues in the field, which indicated that it is not always easy to replicate the moral licensing effect (cf. Chapter 4), which suggests that the effect may not be robust or subject to moderating factors. Thus, a meta-analysis seems important to 1) attain a good indication of the effect size so we know whether real-life interventions are useful, 2) be able to run power analyses to create well-powered studies (and lower the chance of type II errors), and 3) advance the existing theoretical framework through testing for possible boundary conditions of the effect. The purpose of this chapter is therefore to provide a state-of-the-art overview of moral licensing. We report a quantitative meta-analytical review through examining published and unpublished experimental comparisons between a moral licensing and a control condition to determine the magnitude of the moral licensing effect and to identify the potential moderators of the effect.

Moderators of the moral licensing effect

We will explore the conditions under which the moral licensing effect is likely to occur through analyzing theoretically meaningful and methodological moderators. We will examine the following theoretically relevant moderators: the type of moral licensing induction, the behavior measured in the dependent variable, and the domain in which the behaviors take place. We will also examine methodological factors that do not touch upon theoretical distinctions in moral licensing, but are related to the specific

aspects of the research design and the current status of the research: article status and control condition.

Moral licensing induction: traits versus actions

We will compare the effects of moral licensing inductions related to prior moral traits versus prior moral actions. Conway and Peetz (2012) found that recalling moral actions led to licensing (subsequent less moral behavior), whereas recalling moral traits led to consistency (subsequent moral behavior). They argued that the recall of a moral action signals that progress towards the goal of being moral has been made, and for a subsequent choice between doing the morally right thing versus acting out of self-interest (e.g., whether or not to donate money to someone in need), the goal of being moral becomes less important (because one already made progress towards that goal) and self-interest is thus more likely to win. In contrast, recalling moral traits is thought to activate more abstract moral identity concerns. As the recall activates the general concept of morality, people will subsequently behave more according to moral norms (and thus no moral licensing is expected). This theory predicts that licensing effects should only occur when induced through good actions rather than good traits. Study 3 of Conway and Peetz (2012) provides initial support for this idea. In the current meta-analysis we have the opportunity to further test this moderator.

Behavior in the dependent variable: actual versus hypothetical

We will also investigate whether the moral licensing effect differs depending on whether the dependent variable is actual behavior rather than hypothetical behavior (for instance, donating actual money versus indicating the amount of money one would be willing to donate). Previous research revealed that people want to appear moral while avoiding the cost of being moral (Batson & Thompson, 2001; Dana, Weber, & Kuang, 2007). That is, people will display good behavior, as long as the costs of good behavior do not override the benefits of self-interested behavior too much. Since it is relatively effortless to display hypothetical good behavior (talk is cheap), in these cases there may be lower willingness to display undesirable behaviors that one needs to

justify. Thus, one could expect that the moral licensing effect is larger when the dependent behavior consists of actual compared to hypothetical behavior.

Domain: same versus different

We will further investigate whether the size of the moral licensing effect differs depending on whether the good and bad behavior occur in the same or in a different domain. Miller and Effron (2010) pointed out that good behavior in one domain can not only license people to perform dubious behavior in the same domain, but also in unrelated domains (e.g., Mazar & Zhong, 2010). Theory and empirical work in the field of mental accounting reveals that people use different mental accounts to organize their finances (Thaler, 1985). For instance, if people receive a financial windfall in one domain (e.g., a refund for a delayed flight), they typically more easily spend it on something related (e.g., a luxurious dinner during that trip). If similar effects occur in the realm of moral licensing, one might expect that after doing something good in one domain, people more easily allow themselves to do something more questionable in the same domain later. Therefore, it could be expected that licensing effects are larger (smaller) when the moral and immoral behaviors are measured in the same (different) domain.

Article Status: published versus unpublished work

We will examine whether the magnitude of the moral licensing effect depends on whether the study is part of a published article or not. It could be expected that the moral licensing effect, like other empirical findings, is larger for studies in published articles since more positive findings than null- or negative- findings tend to be published (Ioannidis, 2005). We will later test for potential publication bias in various ways, but think it is essential to include study status as a control variable when we examine other possible moderators as well.

Control condition: neutral versus negative

We will further investigate whether the moral licensing effect differs depending on whether the moral licensing condition (in which people are for example asked to recall prior good behavior) is contrasted with a negative control condition (a recall of prior

bad behavior) or with a neutral control condition (a recall of neutral behavior). This is important because the opposite pattern can exist as well: Positive behavior becomes more likely after recalling previous negative behavior, which is called the moral cleansing effect (Conway & Peetz, 2012, Jordan et al., 2011, Sachdeva et al., 2009; Zhong & Liljenquist, 2006, but see also Chapter 4). Based on the theory of moral cleansing, one could expect that the moral licensing effect is larger when a negative control condition is used compared to a neutral one.

In the moral licensing literature, a number of other moderators has been proposed. Power was insufficient to include all these possible moderators, because 1) the current number of included studies (n = 91) does not allow for too many moderators and 2) many of these hypothesized moderators were only tested in very few studies, which would make comparisons unreliable due to the small sample. For completeness of our review of moral licensing effects we do mention the proposed moderators here, hoping that future research can further explore these possible moderators. Specific moderators that were identified (stronger licensing in conditions displayed in italics) are *free* versus forced choice good behavior (Bradley-Geist, King, Skorinko, Hebl, & McKenna, 2010), *high* versus low rationalizability of cheating (Brown et al., 2011), recalling *recent* versus distant good behavior (Conway & Peetz, 2012), having an *outcome-based* versus a rule-based mindset (Cornelissen et al., 2013a), focusing on *goal progress* versus goal commitment (Mann & Kawakami, 2012), and having *no external incentive* versus having an external incentive for one's moral behavior (Clot, Grolleau, & Ibanez, 2013b; Khan & Dhar, 2006).

In addition, some studies tested the moderating effect of continuous personality variables that we could not include in the current meta-analysis. These were being *high* versus low in self-monitoring (Cornelissen, Karelaia, & Soyer, 2013b), having a *high* versus low score on the modern racism scale (Effron et al., 2009; 2012), and having a *strong* versus weak pro-environmental identity when possibly engaging in licensing in the environmental domain (Meijers, Noordewier, Verlegh, & Smit, 2014).

Taken together, we provide a quantitative meta-analytical review of moral licensing through examining both published and unpublished research. We will investigate the

estimated mean effect size of moral licensing and advance the existing theoretical framework on moral licensing through investigating several moderators.

Method

Data collection

An extensive literature search was conducted to collect data on moral licensing, based on the definition of Merritt et al. (2010, p.344) "Past good deeds [or good intentions] liberate individuals to engage in behaviors that are immoral, unethical, or otherwise problematic, behaviors that they would otherwise avoid for fear of feeling or appearing less moral". This included searches in Web of Science and Google Scholar using the following keywords[1]: *(self-)licensing, moral licensing, psychological licensing, moral balancing, moral compensation, moral spillover, self-justification, (moral) credentialing and (moral) credential(s).* Relevant journals (all RSS feeds from the European Journal of Social Psychology, Journal of Consumer Psychology, Journal of Consumer Research, Journal of Experimental Social Psychology, Journal of Marketing Research, Journal of Personality and Social Psychology, Personality and Social Psychology Bulletin, Psychological Bulletin and Psychological Science), conference proceedings, dissertations, and master theses were also checked. Furthermore, we called for relevant studies on moral licensing (both published and unpublished, both successful and unsuccessful) on the Society for Personality and Social Psychology (SPSP) internet forum (10/18/2012) as well as via various mailing lists (SPSP, the Society for Judgment and Decision Making (SJDM), and the Dutch Society of Social Psychologists (ASPO)). Finally, we presented a preliminary version of the meta-analysis at the annual Society for Personality and Social Psychology Conference in New Orleans on January 17th 2013. At this conference, we released another call for data on moral licensing. We considered the studies we found and received until December 1st, 2014 for inclusion.

[1] Variants on these keywords were also used.

Inclusion criteria

Studies were included if they met two criteria. First, the behavior that was measured had to meet our definition of moral licensing. This entails that the behavior has to take place in a moral domain. Licensing studies on self-regulation (e.g., Chiou, Yang, & Wan, 2011; De Witt-Huberts, Evers, & De Ridder, 2012; Fishbach & Dhar, 2005; Mukhopadhyay, Sengupta, & Ramanathan, 2008; Mukhopadhyay & Johar, 2009) do not fall under the definition of *moral* licensing and were thus not included in the current meta-analysis. In addition, this implies that the independent variable should consist of (intended) good *behavior* or the recall thereof. For instance, we included the studies by Mazar and Zhong (2010) where the independent variable consists of buying ecofriendly products in a virtual shopping paradigm, but we did not include the study by Eskine (2012) where the independent variable consists of participants merely being exposed to organic products, since the latter does not entail actual or hypothetical behavior. If there is no prior good deed (or recall thereof), there can also be no moral licensing according to our definition. Following this definition also implies that the dependent variable should measure actual or hypothetical behavior of the participants. Studies where the dependent variable consisted of evaluative judgments, for instance an evaluation of one's morality level (Jordan et al., 2011, Study 1) or the perception of general undesirable behavior (Effron, Monin, & Miller, 2013, Study 2), were therefore not included.

Second, reported statistics had to be adequate to calculate effect sizes. When important statistical information was lacking, authors were contacted for more information. When authors did not respond to the initial request, two reminders were sent. Studies that did not meet our inclusion criteria can be found in Appendix 2.1.

Dependent variables

The dependent variables reported in the included studies comprise immoral behavior, such as cheating and stealing (Mazar & Zhong, 2010), or a decrease in moral behavior, such as donating less money to charity (Sachdeva et al., 2009). These behaviors are real, such as stealing money from the experimenter (Mazar & Zhong, 2010), or hypothetical,

such as indicating that one would be willing to volunteer (Conway & Peetz, 2012). They were either measured on a continuous scale or as a dichotomous choice between a virtuous and a less virtuous option. All effect sizes were recoded so that positive effect sizes indicated associations between previous moral behavior and a subsequent decrease in moral behavior.

Moderators

In several studies that investigated possible moderators of the moral licensing effect, the authors predicted a moral licensing effect in one condition but no moral licensing in another condition. For instance, Conway and Peetz (2012) predicted that licensing would occur when participants recalled a recent moral act, but not when they recalled a temporally distant moral act. In these cases, our default was that we did not include the conditions where the authors did not expect a moral licensing effect. For the Conway and Peetz example on recalling recent versus distant moral acts, this implies that we only included the condition where participants recalled a recent moral act. Across all the studies that we included, we made two exceptions to this default: First, Bradley-Geist et al. (2010) predicted that participants who freely chose to write about a positive (versus negative) experience with a member of a minority group would obtain a moral license, whereas participants who were forced to write about a positive (versus negative) past experience with a member of a minority group would not obtain a moral license. Because in many other studies on moral licensing participants are specifically asked to write about moral behavior or moral traits in the past and thus do not have a choice to write about immoral versus moral behavior (Blanken, Van de Ven, & Zeelenberg, 2012; Chapter 2; Clot et al., 2013b; Conway & Peetz, 2012; Cornelissen et al., 2013a; Jordan et al., 2011; Sachdeva et al., 2009; Schüler, Lehnhardt, & Huber, 2012; Thomas & Showers, 2012; Young, Chakroff, & Tom, 2012), we decided to include the forced choice conditions from Bradley-Geist et al. (2010). Second, in Study 3 of Conway and Peetz (2012) the authors predicted that participants who wrote about moral actions would obtain a moral license, whereas participants who wrote about moral traits would not obtain a moral license. We decided to include both the moral

action and the moral trait conditions because we tested for the difference between moral traits versus actions in our between-study moderator section.

For studies where authors predicted that one specific condition could have a larger licensing effect than the other condition, we included both effects. For instance, Monin and Miller (2001) predicted that participants who previously established nonracist credentials were more likely to display preference for a white over a black person in a hypothetical job hiring task compared to participants who did not establish nonracist credentials. They also predicted that this effect could be larger for participants who completed the independent and dependent variables in front of the same (versus a different) experimenter. For this study, we thus included both the single experimenter audience and the different experimenter audience conditions, because the authors had predicted a licensing effect in both conditions.

Some studies tested the moderating effect of continuous personality variables (Cornelissen et al., 2013b; Effron et al., 2009; 2012; Meijers et al., 2014). For instance, Effron et al. (2009; 2012) measured whether scores on the modern racism scale moderated participants' preferences for white over black persons. In these cases, we included the main effect size without distinguishing between participants who scored high versus low on the measured personality variable because (1) other studies on moral licensing do not differentiate between these specific variables and (2) according to general theorizing on moral licensing the general licensing induction should work for all individuals.

We analyzed the effects of between-study moderators, that is, moderators that we theoretically predicted to influence the magnitude of the moral licensing effect (as pointed out in the introduction section) by means of a regression model.

Overview of analyzed studies

The dataset contains 91 different comparisons between a moral licensing and a control condition, with a total of 7397 participants, reported in 22 published or forthcoming journal articles and 8 unpublished manuscripts. Table 2.1 provides an overview of the included studies.

Table 2.1

Details of all studies included in the meta-analysis (dummy coding for moderators in last columns)

	Author (s)	Year	Study	*N*	*d*	s^2	Licensing manipulation	Dependent variable	A	B	C	D	E
1	Blanken, van de Ven, Zeelenberg, and Meijers	2014	1	64	0.08	0.06	Pp wrote about themselves using positive trait words	Donation to charity	0	1	0	1	0
2	Blanken, van de Ven, Zeelenberg, and Meijers	2014	2	91	0.27	0.04	Pp wrote about themselves using positive words	Cooperative behavior in a hypothetical commons dilemma	0	0	0	1	0
3	Blanken, van de Ven, Zeelenberg, and Meijers	2014	3	567	-0.05	0.01	Pp wrote about themselves using positive trait words	Donation to charity and cooperative behavior in a hypothetical commons dilemma	0	x	0	1	0
4	Blanken, van de Ven, and Zeelenberg	2012	1	86	-0.08	0.05	Pp recalled having good traits	Participants indicated their willingness to fake illness at work to visit a concert in a hypothetical scenario	0	0	0	0	0
5	Blanken, van de Ven, and Zeelenberg	2012	1	86	0.07	0.05	Pp recalled performing good behavior	Willingness to fake illness at work to visit a concert in a hypothetical scenario	1	0	0	0	0
6	Blanken, van de Ven, and Zeelenberg	2012	2	38	0.05	0.11	Pp indicated whether they are planning to donate their organs after death	Everyday Cooperation Scale	1	0	0	0	0
7	Blanken, van de Ven, and Zeelenberg	2012	3	49	0.49	0.08	Pp recalled having good traits	Hypothetical donation to charity	0	0	0	0	0
8	Blanken, van de Ven, and Zeelenberg	2012	3	47	0.03	0.09	Pp recalled performing good behavior	Hypothetical donation to charity	1	0	0	0	0

9	Blanken, van de Ven, and Zeelenberg	2012	4	64	0.30	0.06	Pp recalled having good traits	Willingness to fake illness at work to visit a concert in a hypothetical scenario	0	0	0	0	0
10	Blanken, van de Ven, and Zeelenberg	2012	4	65	0.54	0.06	Pp recalled having good traits	Willingness to fake illness at work to visit a concert in a hypothetical scenario	0	0	0	0	0
11	Blanken, van de Ven, and Zeelenberg	2012	5	57	-0.11	0.07	Pp recalled helping a friend	Willingness to Volunteer Scale	1	0	0	0	0
12	Blanken, van de Ven, and Zeelenberg	2012	5	53	0.12	0.08	Pp recalled helping a stranger	Willingness to Volunteer Scale	1	0	0	0	0
13	Blanken, van de Ven, and Zeelenberg	2012	6	50	0.61	0.10	Pp recalled having good traits and helping a person who is walking on crutches	Giving 'accidentally' overpaid (actual) money back to experimenter (y/n)	x	1	0	0	0
14	Blanken, van de Ven, and Zeelenberg	2012	6	49	0.48	0.10	Pp recalled performing good behaviors and helping a person who is walking on crutches	Giving 'accidentally' overpaid (actual) money back to experimenter (y/n)	1	1	0	0	0
15	Blanken, van de Ven, and Zeelenberg	2012	7	48	0.63	0.09	Pp recalled having good traits	Willingness to pay for luxurious over purposeful goods	0	0	0	0	0
16	Blanken, van de Ven, and Zeelenberg	2012	7	51	0.22	0.08	Pp recalled performing good behavior	Willingness to pay for luxurious over purposeful goods	1	0	0	0	0
17	Blanken, van de Ven, and Zeelenberg	2012	8	94	0.86	0.07	Pp recalled having good traits	Everyday Cooperation Scale	0	0	0	0	0
18	Blanken, van de Ven, and Zeelenberg	2012	8	133	0.24	0.03	Pp recalled performing good behavior	Everyday Cooperation Scale	1	0	0	0	0
19	Blanken, van de Ven, and Zeelenberg	2012	9	65	-0.19	0.06	Pp recalled having good traits	Willingness to pay for luxurious over purposeful goods	0	0	0	0	0

20	Blanken, van de Ven, and Zeelenberg	2012	9	61	0.04	0.07	Pp recalled performing good behavior	Willingness to pay for luxurious over purposeful goods	1	0	0	0	0
21	Blanken, van de Ven, and Zeelenberg	2012	10	61	-0.34	0.07	Pp indicated whether they are planning to donate their organs after death + provided their address for a folder on organ donation	Everyday Cooperation Scale	1	0	0	0	0
22	Blanken, van de Ven, and Zeelenberg	2012	11	83	-0.25	0.05	Pp indicated preference for a black doctor	Willingness to hire white person for a job (over a black person) in a scenario	1	0	1	0	0
23	Blanken, van de Ven, and Zeelenberg	2012	12	57	-0.67	0.08	Pp disagreed with racist statement	Willingness to hire white person for a job (over a black person) in a scenario	1	0	1	0	0
24	Bradley-Geist, King, Skorinko, Hebl, and McKenna	2010	1	38	-0.17	0.11	Pp were asked to write about a positive experience with a black individual	Willingness to hire white person for a job (over a black person) in a scenario	1	0	1	1	0
25	Bradley-Geist, King, Skorinko, Hebl, and McKenna	2010	1	35	0.66	0.13	Pp were given the choice to write about a positive or negative experience with a black individual. Participants who chose to write about a positive experience were included.	Willingness to hire white person for a job (over a black person) in a scenario	1	0	1	1	0
26	Bradley-Geist, King, Skorinko, Hebl, and McKenna	2010	2	44	0.17	0.09	Pp were asked to write about a positive experience with a hispanic individual	Willingness to hire white person for a job (over a hispanic person) in a scenario [2] + tolerance for prejudice scale	1	0	1	1	0
27	Bradley-Geist, King, Skorinko, Hebl, and McKenna	2010	2	42	0.47	0.10	Pp were given the choice to write about a positive or negative experience with a hispanic individual. Participants who chose to write about a positive experience were included.	Willingness to hire white person for a job (over a hispanic person) in a scenario + tolerance for prejudice scale	1	0	1	1	0

[2] Data on this measure were not included since there were insufficient statistics to calculate a cohen's *d* effect size

28	Bradley-Geist, King, Skorinko, Hebl, and McKenna	2010	4	56	-0.70	0.08	Pp were asked to write about a positive experience with a black friend	Willingness to hire white person for a job (over a hispanic person) in a scenario [3] + tolerance for prejudice scale	1	0	1	1	1
29	Brown et al.	2011	1	97	0.41	0.04	Pp read four moral dilemmas and rated how likely they would be to behave in a prosocial fashion if they were to encounter such a situation.	The extent to which participants cheat on a math task	1	1	0	1	0
30	Cain, Loewenstein, and Moore	2005	1	43	0.58	0.10	Pp disclosed their conflict of interest to other pp	Giving worse advice to increase own gains in a hypothetical conflict of interest	1	0	1	1	0
31	Cain, Loewenstein, and Moore	2010	1	348	-0.15	0.01	Pp disclosed their conflict of interest to other pp	Giving worse advice to increase own gains in a hypothetical conflict of interest	1	0	1	1	0
32	Cain, Loewenstein, and Moore	2010	3	49	0.04	0.08	Pp disclosed their conflict of interest to other pp	Giving worse advice to increase own payoff	1	1	1	1	0
33	Cain, Loewenstein, and Moore	2010	3	77	0.40	0.05	Pp disclosed their conflict of interest to other pp	Giving worse advice to increase own payoff	1	1	1	1	0
34	Cascio and Plant	2015	1	86	0.54	0.05	Pp were asked whether they were interested in taking part in a fundraiser (skip a meal for red cross)	Willingness to hire white person for a job (over a black person) in a scenario	1	0	0	1	0
35	Cascio and Plant	2015	2	51	0.66	0.08	Pp were asked whether they were interested in taking part in a fundraiser (skip a meal for red cross)	Willingness to hire white person for a job (over a black person) in a scenario	1	0	0	1	0

[3] Data on this measure were not included since there were insufficient statistics to calculate a cohen's *d* effect size

36	Cascio and Plant	2015	3	54	1.26	0.09	Pp were asked whether they would be willing to donate blood	Explicit racism - Attitudes Towards Black Scale (Example item: "I would rather not have Blacks live in the same apartment building I live in")	1	0	0	1	0
37	Cascio and Plant	2015	4	74	0.60	0.06	Pp were asked whether they were interested in taking part in a fundraiser (skip a meal for red cross)	Indicating stereotypes against blacks	1	0	0	1	0
38	Choi, Crandall, and La	2014	2	116	0.44	0.04	Pp first evaluated a high-quality ad with a Black model and were next given a chance to express positive attitude toward the ad	Evaluation of a second (target) ad including a black model on attractiveness	1	0	1	1	0
39	Choi, Crandall, and La	2014	3	60	0.54	0.07	Pp first evaluated a high-quality ad with a Black model and were next given a chance to express positive attitude toward the ad	Evaluation of a second (target) ad including a black model on attractiveness	1	0	1	1	0
40	Clot, Grolleau, and Ibanez	2014	1	100	0.35	0.04	Pp imagined that they previously performed a good deed	Cheating for higher payoff	1	1	0	1	0
41	Clot, Groleau, and Ibanez	2013a	1	367	0.23	0.01	Pp imagined performing voluntary work	Pp indicated the extent to which they would take actual money out of this public funds	1	1	0	0	0
42	Clot, Grolleau, and Ibanez	2013b	1	192	0.30	0.02	Pp imagined engaging in a pro-environmental activity	Pp could allocate an amount (30€) between themselves and an environmental charity (either World Wildlife Fund or Greenpeace)	0	1	1	1	0
43	Conway and Peetz	2012	1	51	0.33	0.08	Pp recalled moral behavior they performed during the previous week (recently)	Willingness to Volunteer Scale	1	0	0	1	1

44	Conway and Peetz	2012	2	39	0.58	0.11	Pp described moral behavior in the recent past (one week ago)	Willingness to Volunteer scale	1	0	0	1	1
45	Conway and Peetz	2012	3	65	0.00	0.06	Pp were asked to visualize having certain moral characteristics, and wrote about what having these traits would mean for their personality.	Donation to charity; pp entered a draw and they could indicate how much of the price money would be donated to charity	0	1	0	1	1
46	Conway and Peetz	2012	3	65	0.79	0.07	Pp were asked to visualize performing moral behaviors, and wrote about how they would perform these behaviors.	Donation to charity; pp entered a draw and they could indicate how much of the price money would be donated to charity	1	1	0	1	1
47	Cornelissen, Bashshur, Rode, and Le Menestrel	2013	1	48	0.59	0.09	Pp remembered an episode in the past where they did something ethical	Amount of money offered in actual dictator game	1	1	0	1	1
48	Cornelissen, Bashshur, Rode, and Le Menestrel	2013	2	40	0.84	0.11	Pp remembered an episode in the past where they did something ethical	Amount of money offered in actual dictator game	1	1	0	1	1
49	Cornelissen, Bashshur, Rode, and Le Menestrel	2013	3	50	0.53	0.08	Pp remembered an episode in the past where they did something ethical	Cheating for higher payoff	1	1	0	1	1
50	Cornelissen, Karelaia, Soyer	2013	1	70	0.57	0.06	Pp could indicate whether they supported UNICEF.	Providing help for organization by investing time to develop slogans.	1	1	1	0	1
51	Cornelissen, Karelaia, Soyer	2013	2	92	0.36	0.04	Pp indicated whether they supported Oxfam.	Contribute part of participation fee to Oxfam.	1	1	1	0	1
52	Effron, Cameron, and Monin	2009	1	84	0.44	0.05	Pp indicated that they would vote for president Obama (displaying preference for black over white president)	Willingness to hire white person for a job (over a black person) in a scenario	1	0	1	1	0

53	Effron, Cameron, and Monin	2009	2	40	0.35	0.10	Pp indicated that they would vote for president Obama (displaying preference for black over white president)	Willingness to hire white person for a job (over a black person) in a scenario	1	0	1	1	0
54	Effron, Cameron, and Monin	2009	3	50	0.16	0.08	Pp indicated that they would vote for president Obama (displaying preference for black over white president)	Willingness to allocate funds to a group of white (over black) people in a scenario	1	0	1	1	0
55	Effron, Monin, and Miller	2012	3	157	0.25	0.03	Pp displayed nonracist behavior while having a racist alternative to their behavior (accused white over black guy for crime)	Willingness to hire white person for a job (over a black person) in a scenario and willingness to allocate funds to a group of white (over black) people	1	0	1	1	0
56	Jordan, Mullen and Murnighan	2011	2	68	0.42	0.06	Pp recalled a situation in which they helped other people	Indicate willingness to donate money to charity, donate blood, volunteer	1	0	0	1	0
57	Jordan, Mullen and Murnighan	2011	3	76	0.98	0.06	Pp recalled a situation in which they helped other people	The extent to which participants cheat on math task	1	1	0	1	0
58	Khan and Dhar	2006	1	108	0.62	0.04	Pp imagined that they had volunteered to perform community service	Preferring a vicious over a less vicious item	1	0	0	1	0
59	Khan and Dhar	2006	2a	94	0.60	0.05	Pp imagined donating money to charity	Preferring a vicious over a less vicious item	1	0	0	1	0
60	Khan and Dhar	2006	2b	80	0.56	0.06	Pp imagined donating money to charity	Preferring a vicious over a less vicious item	1	0	0	1	0
61	Khan and Dhar	2006	3	80	0.39	0.05	Pp imagined helping a foreign student	Pp received $2 and could indicate whether and how much money they would donate to charity	1	1	0	1	0

62	Khan and Dhar	2006	4	80	0.46	0.06	Pp imagined that they had volunteered to perform community service	Preferring a vicious over a less vicious item	1	0	0	1	0
63	Khan and Dhar	2006	5	66	0.56	0.06	Pp imagined that they had volunteered to perform community service	Preferring a vicious over a less vicious item	1	0	0	1	0
64	Leonard	2012	1	32	-0.29	0.13	Pp imagined refusing an offer to fake volunteering less hours than required	Indicate how much money they would keep from a wallet they found in a hypothetical scenario	1	0	0	0	0
65	Leonard	2012	1	92	-0.04	0.04	Pp imagined refusing an offer to fake volunteering less hours than required	Indicate how much money they would keep from a wallet they found in a hypothetical scenario	1	0	0	0	1
66	Leonard	2012	2	43	-0.02	0.09	Pp imagined refusing an offer to fake volunteering less hours than required	Indicate how much money they would keep from a wallet they found in a hypothetical scenario	1	0	0	0	1
67	Leonard	2012	3	64	0.01	0.06	Pp imagined that they found a wallet on the floor and decided not to take any of the money	Pp indicated to what extent they would refuse an offer to fake volunteering less hours than required	1	0	0	0	1
68	Mann and Kawakami	2012	1	44	0.73	0.10	Pp received feedback that they were drawing closer to their goal (try to have positive evaluations of black people whenever they were presented with an image of blacks)	Interpersonal closeness task – seating distance towards black person	1	1	1	1	1
69	Mann and Kawakami	2012	2	94	0.55	0.04	Pp received feedback that they were drawing closer to their goal (try to have positive evaluations of black people whenever they were presented with an image of blacks	Racial attitudes	1	0	1	1	1

70	Mann and Kawakami	2012	3	30	0.86	0.15	Pp received visual feedback during a task that indicated that they were becoming more positive toward Blacks based on physiological responses measured with a LifeShirt	Interpersonal closeness task – seating distance towards black person	1	1	1	1	1
71	Mazar and Zhong	2010	2	76	0.29	0.05	Pp were assigned to an on-line store with more green (ecofriendly) than conventional products. Pp made purchases in this store	Amount of money offered in an actual dictator game	1	1	0	1	0
72	Mazar and Zhong	2010	3	81	0.53	0.05	Pp were assigned to an on-line store with more green (ecofriendly) than conventional products. Pp made purchases in this store	Lying about performance to gain extra money for the experiment and stealing money from envelope	1	1	0	1	0
73	Meijers, Noordewier, Verlegh, and Smit	2014	1	40	-0.04	0.10	Pp imagined buying environmentally friendly sneakers.	Sustainable intentions measured through the Minton and Rose Behavioral Intentions Scale	1	0	1	0	0
74	Meijers, Noordewier, Verlegh, and Smit	2014	1	40	0.83	0.11	Pp imagined buying environmentally friendly sneakers.	Sustainable intentions measured through the Minton and Rose Behavioral Intentions Scale	1	0	1	0	0
75	Meijers, Noordewier, Verlegh, and Smit	2014	2	88	0.60	0.05	Pp entered a webshop with heavily environmental apparel and were asked to compose an outfit.	Environmental Concern & Pro-environmental intentions	1	0	1	0	0
76	Meijers, Noordewier, Verlegh, and Smit	2014	2	88	-0.24	0.05	Pp entered a webshop with heavily environmental apparel and were asked to compose an outfit.	Environmental Concern & Pro-environmental intentions	1	0	1	0	0

77	Merritt, Effron, Fein, Savitsky, Tuller, and Monin	2012	2	70	0.48	0.06	Pp had a chance to establish credentials by identifying ambiguous behaviors as racist	Willingness to hire white person for a job (over a black person) in a scenario	1	0	1	1	0
78	Monin and Miller	2001	1	140	0.26	0.03	Pp indicated whether they considered each of five negative statements about women to be right or wrong, for instance: "Most women are not smart"	Willingness to hire a woman for a job (over a man) in a scenario	1	0	1	1	0
79	Monin and Miller	2001	2	110	0.71	0.04	Pp hired a women (showing that they are not sexist) or a black guy (showing that they are not racist)	Willingness to hire white person for a job (over a black person) in a scenario	1	0	1	1	0
80	Monin and Miller	2001	3	21	0.91	0.21	Pp hired a black guy (showing that they are not racist)	Willingness to hire white person for a job (over a black person) in a scenario	1	0	1	1	0
81	Monin and Miller	2001	3	19	0.89	0.23	Pp hired a black guy (showing that they are not racist)	Willingness to hire white person for a job (over a black person) in a scenario	1	0	1	1	0
82	Sachdeva, Iliev, and Medin	2009	1	29	0.60	0.14	Pp wrote about themselves using positive words	Donation to charity	0	1	0	1	0
83	Sachdeva, Iliev, and Medin	2009	2	18	1.10	0.26	Pp wrote about themselves using positive words	Donation to charity	0	1	0	1	1
84	Sachdeva, Iliev, and Medin	2009	3	31	0.57	0.13	Pp wrote about themselves using positive words	Cooperative behavior in a hypothetical commons dilemma	0	0	0	1	0
85	Schüler, Lehnhardt, and Huber	2012	1	62	-0.42	0.07	Pp recalled a situation in which they helped other people	Dictator game where pp could earn points for lottery	1	1	0	0	0
86	Spektor	2014	1	278	0.00	0.01	Pp remembered and wrote down a concrete situation within the last 7 days in which they felt righteous or honorable (moral condition)	Pp were told that they automatically participated in a €100 draw and could donate an amount of their choice to one or more charities in case they won	1	1	0	0	1

87	Susewind and Hoelzl	2014	1	84	0.38	0.05	Pp focused on progress towards sustainable goals	Pp indicated their prosocial intentions	1	1	0	1	0
88	Susewind and Hoelzl	2014	2	62	0.44	0.07	Pp brainstormed on a topic that clearly benefited others and made progress on their goals	Pp divided €6 between themselves and another participant of the study that was randomly assigned to them, knowing that the other participant has to accept their decision	1	1	0	1	1
89	Thomas and Showers	2012	2	76	-0.30	0.05	Pp recalled a time when they did something especially moral or ethical.	The extent to which participants cheat on a math task	1	1	0	0	1
90	Young, Chakroff, and Tom	2012	1	66	0.41	0.06	Pps were asked to describe 5 good deeds	Donation amount to charity - hypothetical amount, actual y/n and actual amount	1	x	0	1	0
91	Zhong, Ku, Lount and Murnighan	2009	2	68	0.29	0.06	Pp were asked to imagine that they would display ethical behavior in a specific situation	Willingness to display ethical behavior in hypothetical situation	1	0	1	1	0

Note. N = Number of participants included in study; *d* = Cohen's d effect size; s^2 = Within-study variance; A = Licensing induction: Traits (0) versus actions (1); B = Behavior measured in the dependent variable: Actual (1) versus hypothetical (0); C = Domain: Same (1) versus different (0); D = Article status: Published (1) versus unpublished (0); E = Control condition: Neutral (0) versus negative (1).

Effect size measure

We calculated Cohen's *d* based on pooled standard deviations for all studies, with a positive *d*-value indicating moral licensing. For calculation of the effect sizes of the continuous dependent variables we used means and standard deviations. When information on means and standard deviations was lacking, *t*-values were used for the calculation. For calculation of the effect sizes of dichotomous dependent variables we used the reported χ^2 statistic or calculated the χ^2 statistic manually. For all effect sizes, we applied the small sample bias correction provided by Lipsey and Wilson (2001). The equations that we used to calculate the effect sizes can be found in Appendix 2.2.

When the moral licensing condition (e.g., recall positive behavior) was compared to both a neutral (e.g., recall neutral behavior) and a negative (e.g., recall bad behavior) control condition, we always report the comparison between the moral licensing condition and the neutral control condition. Differences between comparisons with negative conditions versus comparisons with control conditions are analyzed in the between-moderator section.

Several studies reported the effect of one independent variable on multiple dependent variables. For instance, Jordan et al. (2011) measured the effect of recalling (un)ethical behavior on (1) allowing an answer to a math task to appear on the screen, (2) whether the participant used that provided answer, and (3) the number of answers completed before participants started to cheat. To avoid statistical dependencies, in these situations we combined the different effects into a single effect size by averaging the multiple effect sizes (Hedges & Olkin, 1985; Marascuillo, Busk & Serlin, 1988; Rosenthal & Rubin, 1986). Some studies included two moral licensing conditions. For instance, Blanken et al. (2012) measured the effects of displaying prior good behavior and of displaying having been a good person on willingness to help and compared both conditions to a similar control condition. In these cases we report a separate effect size for each moral licensing condition. This also enabled us to code these conditions independently for the moderator section.

Random-effects model

We examined the overall effect size of the moral licensing effect using a random effects model, since there was no a priori reason to assume that the true effect size is exactly the same across all studies (Hedges & Vevea, 1998). The model treats the heterogeneity as purely random; $\theta_i = \mu + u_i$, where $u_i \sim N\ (0, \tau^2)$ and τ^2 is the residual heterogenity estimated with the DerSimonian-Laird estimator (DerSimonian & Laird, 1986). The average true effect μ is calculated via weighted least squares with weights equal to $w_i = 1/(v_i + \hat{\tau}^2)$ where $\hat{\tau}^2$ is the estimate of τ^2 and v_i is the sample variance (Viechtbauer, 2010).

Results[4]

Mean effect of moral licensing

The random effects meta-analysis (N = 91; 57 published and 34 unpublished studies) produced a mean effect size of moral licensing of d = 0.31 ([CI_{95}] 0.23 to 0.38). The null hypothesis H0: μ = 0 was rejected (Z = 8.24, p < .001), showing that there is a significant moral licensing effect across the studies we analyzed. The between-study variance is τ^2 = .06 ([CI_{95}] .03 to .11) with I^2 = 54.58% of the total variation due to heterogeneity among true effects; in other words, 54.58% of the variability may be attributable to systematic between-study differences (Higgins & Thompson, 2002). Also, the test for heterogeneity is significant $Q(df = 90) = 198.17$, p < .001, implying that other possible moderators are influencing the magnitude of the moral licensing effect. The forest plot of the meta-analysis is depicted in Figure 2.1. To conclude the main analysis, the moral licensing effect is small to medium in effect size. In addition, there is substantial variation of the effect size between studies. In the next section, we discuss possible publication bias and whether specific between-study moderating variables are responsible for this variation.

[4] All analyses were carried out with the statistical software R, using the metafor package (Viechtbauer, 2010).

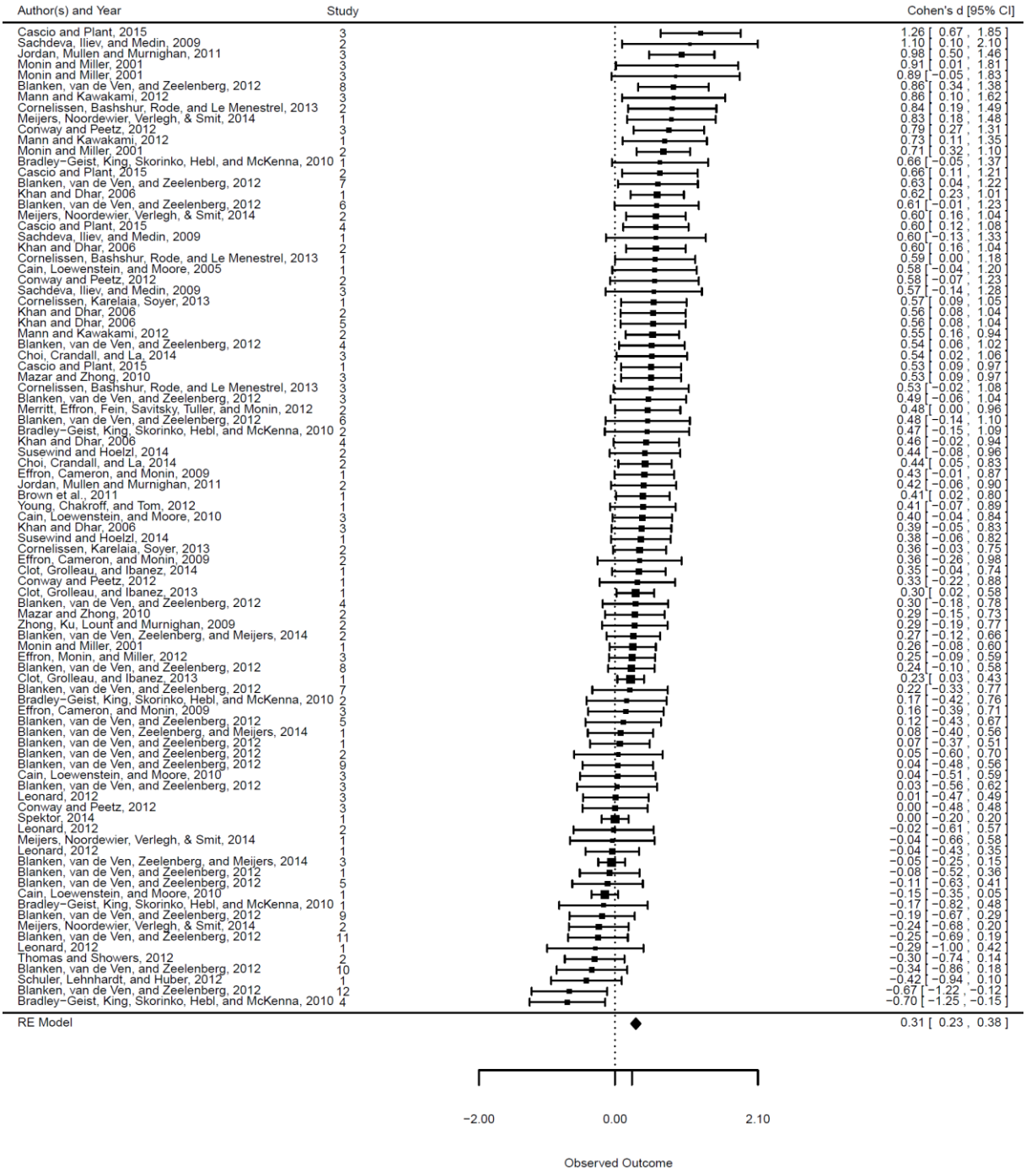


Figure 2.1. Forest plot of all included studies on moral licensing. RE Model = Random Effects Model, Observed Outcome = The obtained average effect size. This figure was created using the metafor forest (res) function in the R metaphor package.

The impact of publication bias

The existence of a positive publication bias in the licensing literature was assessed via two methods; by creating a funnel plot and by investigating article status (published versus unpublished) as a between-study moderator. First, we created a funnel plot of the effects sizes of the published data (n = 57) against their corresponding standard errors. If there is no positive publication bias, the funnel plot should be roughly symmetrical around the true effect size estimate, because without a publication bias an equal amount of studies should find smaller effects than the true effect size as there are studies that find larger effects. If there is a positive publication bias, high powered studies should be close to the average effect size and be present on both sides, whereas studies with smaller samples and higher variability would be more likely to only appear as being larger than the true effect size. Smaller studies that find no significant effect are unlikely to be published, and therefore the bottom side of the funnel plot remains relatively empty in the presence of a positive publication bias.

Figure 2.2 shows the effect sizes in the random effects funnel plot with the filled-in data. In this plot, the effect size estimates from the included published studies are represented as black dots. The white dots represent the estimated number of missing studies (n = 21). Interpretation of the symmetry of the black dots in the funnel plot is rather subjective (Thornton & Lee, 2000), but a regression analysis can be conducted with the standard error as predictor of the observed outcomes. When there is a publication bias, the observed effect sizes are positively related to the standard error (showing that studies with smaller sample sizes find larger effects; Egger, Smith, Schneider, & Minder, 1997). The regression test for funnel plot asymmetry using the weighted regression with multiplicative dispersion model showed a significant effect, $t(55) = 5.72$, $p < .001$, indicating the presence of a positive publication bias.

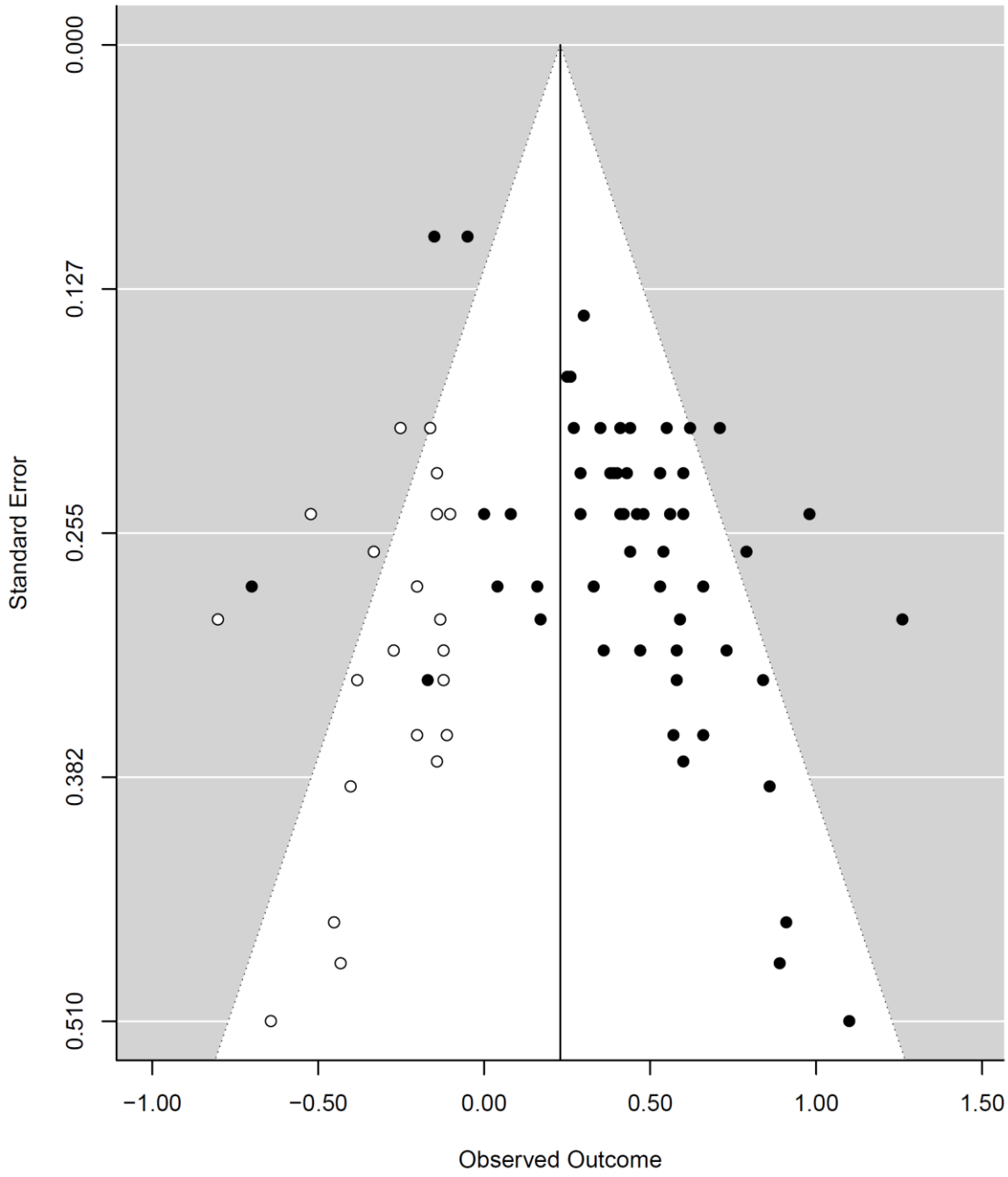


Figure 2.2. Trimmed and filled funnel plot of all published studies (n = 57) on moral licensing. The black dots represent the published studies on moral licensing. The white dots represent the estimated missing studies (n = 21). This figure was created using the metafor funnel (rtf) function in the R metafor package.

Second, we tested for study status (published versus unpublished) as a between-study moderator. Article status (published versus unpublished) turned out to significantly influence the estimated moral licensing effect size, with published studies having larger effects (d = 0.43, SE = 0.04) than unpublished studies (d = 0.11, SE = 0.06), Q_M[5] (1) = 20.66, p < .001. Both methods thus indicate the existence of a publication bias. We decided to keep the article status moderator in the overall meta-regression model to control for its effects because other between-study moderators might be mainly tested in published studies.

Between-study moderators

To estimate the amount of variance in effect sizes of the included studies that could be explained by the theoretically predicted between-study moderators, we fitted a meta-regression model including all these moderators in the R script (Viechtbauer, 2010). Two independent coders coded the potential between-study moderators (see also the last column of Table 2.1). When there was disagreement, the two coders talked to the primary researcher and came to a joint conclusion. For three studies (Blanken et al., 2012; 2014; Young et al, 2012), it was not possible to code one of the moderators. For instance, the effect size of Young et al. (2012) consisted of merged dependent variables. Since one variable measured hypothetical donation behavior and another variable measured actual donation behavior, it was not possible to indicate whether the behavior was actual or hypothetical. We excluded these three studies, leaving n = 88 studies for the moderator analyses. The initial inter-rater reliability for each moderator was as follows: Independent variable: Trait versus Action Cohen's κ = .97; Dependent variable: Actual versus Hypothetical behavior Cohen's κ = .92; Domain: Same versus Different Cohen's κ = .64; Control condition: Neutral versus Negative Cohen's κ = 1.00. Initially, the moderator 'Dependent variable: A decrease in good behavior versus an increase in bad behavior' was also coded. Since coding this moderator seemed problematic for both independent coders (for instance, is a preference for a white over

[5] Qm is an omnibus test that analyzes whether the effect sizes for the two moderator conditions significantly differ (Zhang, 1999).

a black job applicant an increase in bad behavior versus a decrease in good behavior), we decided not to include this moderator in the main regression model.

First, we conducted separate analyses for the different moderators. The type of independent variable (trait versus action) did not influence the average effect size, $Q_M(1) = 0.05$, p = .819. The type of behavior measured in the dependent variable (actual versus hypothetical) did not influence the average effect size, $Q_M(1) = 0.86$, p = .354. Domain (same versus different) did not influence the average effect size, $Q_M(1) = 0.18$, p = .670. As mentioned in the publication bias section, article status (published versus unpublished) did significantly influence the estimated moral licensing effect size. The type of control condition (neutral versus negative) did not influence the average effect size, $Q_M(1) = 0.03$, p = .860. Next, we analyzed the moderators together in a multiple regression model to control for collinearity. Table 2.2 provides an overview of the estimated β coefficients for each between-study moderator and the *z*-scores, standard errors, and *p*-values. Also in this multiple regression model the results showed that studies in published papers tend to have larger effect sizes than studies from unpublished work, β = .34, p < .001. Aside from this moderator, no other moderators were found to have a significant effect in the model. Thus, type of independent variable (trait or action), type of dependent variable (actual or hypothetical behavior), domain of dependent variable (same versus different), and type of control condition (neutral versus negative), did not moderate the moral licensing effect size.

Table 2.2

Between-study moderator estimates in the meta-regression

Moderator	**β**	***se***	***Z***	***p***
Intercept	.48	.09	5.30	<.001
Induction:	-.04	.10	-0.39	.696
Actions (d = 0.31) vs traits (d = 0.33)				
Dependent variable:	.02	.08	0.27	.791
Actual (d = 0.36) vs hypothetical behavior (d = 0.28)				
Domain:	-.10	.08	-1.32	.186
Same (d = 0.29) vs different (d = 0.32)				
Status:	.34	.08	4.57	<.001
Published (d = 0.43) vs unpublished (d = 0.11)				
Control condition:	-.01	.09	-0.12	.905
Negative (d = 0.30) vs neutral (d = 0.31)				

Discussion

In the current chapter we aimed to give a state-of-the-art overview of moral licensing by examining the magnitude of the moral licensing effect and testing for potential moderators. We found an overall average effect size of d = 0.31, [CI_{95}] 0.23 to 0.38, that is statistically different from zero, suggesting that there is a small-to-medium moral licensing effect (Cohen, 1992). To contextualize what is meant by small-to-medium, we compared this effect size to other relevant effect sizes in the field. Social psychological effects typically yield a value of r equal to .21 (approx. 4% variance explained) (Richard, Bond, & Stokes-Zoota, 2003). This translates to a Cohen's d value of 0.43. The moral licensing effect size is thus slightly smaller than the average effect size in social psychology. Of course, this does not imply that the moral licensing effect has little theoretical or practical relevance.

An important consequence of this small-to-medium effect size is that properly powered studies on moral licensing need far more participants than are typically used. We used G*power (Faul, Erdfelder, Lang, & Buchner, 2007) to determine that one would need 165 participants per condition to have 80% statistical power to find an effect of $d = 0.31$.[6] A post hoc power analysis on all studies using our current effect size estimate finds that on average, the studies only have 28% power. Increasing sample size in moral licensing studies will help the researcher, as it increases the chance of finding an effect. Furthermore, it will allow for more solid conclusions, for instance about the various moderators that have been tested in different studies.

Potential moderators and underlying mechanisms

In the following section we discuss the moderators that we tested for in our meta-analysis and relate the findings to previously proposed mechanisms of moral licensing. Note that we also tested for and found a publication bias in our moderator tests, and we will return to that topic in a later section.

Moral licensing induction: traits versus actions

Conway and Peetz (2012) hypothesized that recalling prior good actions would lead to moral licensing, while recalling prior good traits would lead to consistency. They found initial support for this hypothesis. Our meta-analysis could not confirm this finding. That is, we found no difference in size of the moral licensing effect, based on whether the prior good deed was coded as an action or a trait. For our moderator section, we included both recalled actions and performed actions (such as buying ecofriendly products; Mazar & Zhong, 2010). Conway and Peetz found that recalled moral actions elicited self-licensed behaviors, whereas recalled moral traits provoked consistent good behaviors (Study 3). It may be the case that the effect obtained by Conway and Peetz is specific to recall paradigms, although we have no theoretical idea

[6] This power analysis is based on a two-sided test. We think this is appropriate, because the alternative hypothesis (consistency in behavior) would also be theoretically likely.

why that would be the case. Future research could further test whether and how this moderator is important for moral licensing to occur.

Behavior in the dependent variable: actual versus hypothetical

It is cheap and easy for people to display hypothetical behavior (i.e., to state good intentions) that is in accordance with their previous laudable behavior, which would make a consistency effect more likely. Therefore, we expected that perhaps moral licensing would be stronger for actual good behavior compared to hypothetical good behavior. However, the meta-analysis showed no such differences between actual versus hypothetical behavior. There was only one study in the dataset that directly investigated similar actual versus hypothetical behavior as a dependent measure in the licensing paradigm. Young et al. (2012) investigated the effects of recalling good (vs bad or neutral) deeds on hypothetical *and* actual donations to charity. They found a consistency effect, such that people who recalled prior good deeds donated nearly twice as much to charity compared to people who recalled prior bad deeds or who recalled a neutral conversation. They did not find a licensing effect. They also did not find any differences between reported hypothetical and actual donations, similar to the results of our meta-analysis. However, note that the actual donation request in that study was directly placed after the hypothetical donation item, which might have facilitated this effect.

Maybe, in the case of a hypothetical dependent variable, people are less tempted to display immoral behavior. When confronted with an actual choice with real consequences, decisions may be different. In these so-called affectively rich states (Rottenstreich, & Hsee, 2001), people are more likely to focus predominantly on the favorability of the outcomes. Therefore, more research is needed that not only focuses on whether the behavior is hypothetical or not, but also on how tempting it is.

Moreover, we think that more careful consideration of the dependent variables being used in moral licensing studies is essential. For example, quite some research on moral licensing includes scales like the Everyday Cooperation Scale (De Hooge, Zeelenberg, & Breugelmans, 2007) and the Willingness to Volunteer Scale (DeVoe & Pfeffer, 2007)

as the dependent measures. Such scales measure people's stated intention to want to help out (and perform good behavior). However, it could be the case that people overstate their willingness to help others, perhaps partly due to social desirability concerns. If a moral licensing condition causes people to indicate a lower willingness to help others, this could indeed reflect licensing. However, it could also imply that people who recall prior good behavior actually become more honest. When they are asked whether they want to volunteer, they might be more thoroughly thinking about it and answering it truthfully, which is usually that one actually does not have the time to volunteer. Both such a consistency process and moral licensing would predict lower scores on for example the Willingness to Volunteer Scale, but the underlying process is completely different. Researchers should be aware of these possible confounds when studying licensing with hypothetical behavioral measures.

Domain: same versus different domain

If effects similar to mental accounting (Thaler, 1985) occur in the realm of moral licensing, licensing effects would be larger when the moral and immoral behaviors were measured in the same domain compared to in a different domain. However, the meta-analysis did not find differences between immoral behaviors in the same domain versus in a different domain. Effron and Monin (2010) reasoned that the relative effectiveness of same-domain versus different-domain moral licensing depends on whether the immoral behavior being licensed is blatantly bad or only ambiguously bad. They tested this idea in studies of observers' inclination to license the behaviors of other persons and found that different-domain licensing worked better than same-domain licensing when the bad deeds were blatant. In contrast, same-domain licensing worked better than different-domain licensing when the bad deeds were ambiguous (i.e., could be construed as non-problematic). Our meta-analysis could not control for the ambiguity of the licensed behavior. We could not make inferences about how blatantly bad the behavior in the dependent variables actually was for participants, because it was not clear to what extent they actually justified the licensed behaviors. Effron and Monin tested their predictions for observer-licensing, but not for self-licensing. Future research on self-licensing could therefore clarify whether the

ambiguity of the licensed behavior plays a role in the distinction between same- versus different-domain licensing.

Control condition: neutral or negative control condition

We were surprised by the absence of a difference in whether the licensing condition (e.g., recall a previous moral action) was compared to a *neutral* control condition (e.g., recall a previous trip to the shopping mall) or a *negative* control condition (e.g., recall a previous immoral action). This finding is surprising because researchers have documented the moral cleansing effect, the finding that recalling previous immoral behavior leads to more subsequent moral behavior (the exact opposite of the moral licensing effect; see Conway & Peetz, 2012; Jordan et al., 2011; Sachdeva et al., 2009; Zhong & Liljenquist, 2006). With the existence of a moral cleansing effect, one would expect that the moral licensing effect should be larger when the positive condition is compared to the negative condition than a neutral condition. After all, if recalling immoral behavior leads to more moral behavior compared to a neutral control condition, the contrast with the recalling moral behavior condition should be even stronger.

There is some discussion about whether the moral cleansing effect is indeed as strong as initial research suggested (cf. Chapter 4). However, another possibility is that perhaps recalling or performing bad or immoral behavior can induce people to feel good about themselves. Research on the ease of retrieval explanation of the availability heuristic shows that reminders of behavior only elicit feelings that are congruent with these behaviors if it is easy for participants to retrieve these behaviors (Schwarz et al., 1991). If participants find it difficult to recall immoral behavior (or can only come up with an instance that was a very long time ago), they might infer from this that they are quite moral persons themselves. Furthermore, exposure to extreme stimuli from a category can also sometimes remind people of the other extreme of this category (Herr, Sherman, & Fazio, 1983). By activating the concept of immoral behavior, the manipulations might also have activated the concept of morality. For these reasons, some of the manipulations that induced immoral behavior might have inadvertently

also activated moral behavior somewhat, thereby leading to a moral licensing effect that is as large as moral licensing contrasted to neutral behaviors.

Taken together, none of the examined moderators that were based on existing theorizing on moral licensing significantly moderated the moral licensing effect in the current meta-analysis. There are various possible explanations for this, in addition to the ones described above. Moderators itself might again have moderators for when the effect occurs or not. The studies that were included in the meta-analysis employed very different designs and included different populations. Variations in experimental design may have resulted in very different outcomes with respect to the moderating variables. In addition, small sample sizes of several included studies may have made it more difficult to perform a proper assessment of the predicted moderators as that made effect size estimates less precise. We therefore recommend researchers to critically revise the research paradigms and increase sample sizes that they employ to study the moral licensing effect and, if necessary, revise the theory.

Other possible mechanisms

The literature on moral licensing does not lack explanatory mechanisms. Conway and Peetz (2012), for instance, proposed that recalling *distant* moral behavior makes one's moral values salient and induces people to act consistent with these moral values, whereas recalling *recent* moral behavior leads to moral licensing. The reason is that recent moral behavior reminds people that they made goal-progress towards their goal of being a good person. Having satisfied that goal, people can then focus attention on other goals, which might include for example selfish behavior. Conway and Peetz indeed found that recalling temporally distant moral behavior led to subsequent good behavior (consistency), whereas recalling recent moral behavior decreased people's tendency to engage in good behavior (licensing). Consistent with that idea, Fishbach and Dhar (2005) found that inducing a sense of achieved goal progress can initiate behaviors associated with licensing in the self-regulation domain such as making unhealthy food choices and spending less time on course work. However, it should be noted that in other studies that find a moral licensing effect, people recall past moral behavior in which it is not specified that it has to be recent or distant moral behavior

(e.g., Jordan et al., 2011; Sachdeva et al., 2009). Therefore, we did not have sufficient information to be able to include the recency of the prior positive behavior as a possible moderator for the moral licensing effect in our meta-analysis.

Another possible mechanism for moral licensing was proposed by Miller and Effron (2010) who distinguished two different pathways of moral licensing, namely (1) moral licensing via credits and (2) moral licensing via credentials. The first pathway, moral licensing via *credits*, asserts that engaging in good behaviors endows people with credits that function to balance out subsequent questionable behaviors, like some sort of moral currency. A person can perceive his recent voluntary work as a credit license to decline a donation request. Thus, a person who obtained a credit might think 'I have done something good so I can now do something bad'. Importantly, the perceived meaning of undesirable acts does not change, but the expending of one's previously earned credit "nullifies" the current bad deed, which makes it acceptable. The deed is still seen as negative, but acceptable. The undesirable behavior is thus accepted, because it is offset by prior good behavior. The second pathway, moral licensing via *credentials*, entails that one's behavioral history provides a license by changing the way subsequent behavior is construed. These credentials function like a character witness on which one can repeatedly call to testify that subsequent dubious behavior is not wrong or immoral. For instance, previous unprejudiced behavior establishes oneself as an unbiased person. A subsequent discriminatory attitude is thought of as less prejudiced, because it came from an ostensibly unbiased source. This means that a bad deed is seen as less bad if a prior good deed preceded it. The credits versus credentials explanation of Miller and Effron could not be tested in our meta-analysis, as it was not possible to code participants' thoughts on this distinction.

Publication bias

A meta-analysis allows one to examine publication bias, which is the tendency that significant results are more likely to get published than non-significant results. The regression model including the between-study moderators revealed that moral licensing studies in published papers tend to have larger effect sizes than studies that did not appear in published papers. The funnel plot in Figure 2.2 also points to

publication bias. As with any publication bias, it can be caused by researchers only submitting positive results, by the tendency of journals to reject studies with negative results, and also by the design or the execution of specific studies (for example, a part of the unpublished papers may suffer from methodological flaws, see Thornton & Lee, 2000). We hope that our estimate of the effect size of moral licensing helps other researchers to come up with well-powered tests, for which it should be easier to publish those also if there is no effect. Running studies with larger samples would thus not only make the effect size estimates more precise, it would hopefully also help to reduce the publication bias.

Possible limitations

Although we believe that the current meta-analysis offers important insights for the moral licensing literature, several limitations of the analysis are noteworthy. First, a substantial amount of the included effect sizes is based on small sample sizes, which could lead to an overestimation of the true effect size (Reynolds & Day, 1984). Another point is that since the moderator analyses are performed on 88 studies, we could select a limited number of moderator variables. Furthermore, as with all meta-analyses, the inclusion criteria are subjective. Although we formulated clear inclusion criteria in advance, the inclusion of some studies was debatable. For instance, we included both studies in which the moral licensing induction exists of (the recall of) actual good behavior as well as studies in which the moral licensing induction focuses on having good intentions. One could argue that both types of independent variables are different in nature and should therefore be analyzed separately. However, according to the current theorizing on moral licensing, both good behaviors and good intentions should lead to the behaviors associated with moral licensing (Khan & Dhar, 2006). Another example is Study 1 by Gneezy, Imas, Brown, Nelson, and Norton (2012), in which the independent variable existed of a donation to charity that was automatically deducted from the participants' payment or a donation to charity that was made by the experimenter on behalf of the participants. Gneezy et al. stated that the donation that was made by the experimenter on behalf of the participants would create a self-license. However, we did not include these studies in our meta-analysis, because we

do not think participants necessarily perceive this kind of behavior as having done something good themselves, and therefore the studies did not meet our definition of moral licensing. Finally, although we explicitly searched for unpublished studies in an attempt to prevent file drawer effects, there are likely studies on moral licensing that we could not detect.

Important aspects for future studies on moral-licensing

Hofmann, Wineski, Brandt, and Skitka (2014) recently investigated everyday morality outside of the lab, through assessing moral and immoral acts in a large community sample (N = 1252).[7] They found a moral licensing pattern, in that people who committed a moral act had a larger likelihood of committing an immoral act later that day. Thus, although the average effect size of moral licensing in experimental paradigms is small-to-medium, moral licensing seems to occur in daily life. The moral licensing effect and the way it is studied thus deserve further attention. In the remainder of this section, we will outline recommendations for scholars studying moral licensing.

The main advice is for researchers studying moral licensing to increase the power of the studies. As explained before, this is important because it helps the researcher by lowering the chance of a study not finding an effect. Note that with the average power of current moral licensing studies (28%), there is a 72% chance to find a non-significant moral licensing effect if there is an effect. Our power calculation reveals that one needs 165 participants per cell, in order to have 80% power to find an effect as large as the one established in our meta-analysis.

Moral licensing is typically seen in the temporal pattern that people who behaved in a good or moral way later feel justified to refrain from socially desirable or morally laudable actions (Miller & Effron, 2010; Monin & Miller, 2001). All studies on moral licensing that we are aware of investigated the phenomenon in terms of two

[7] We did not include this study in our meta-analysis because we wanted to include effect sizes based on the comparison between an experimental licensing condition and a control condition. In this specific study there was no control condition and no random assignment of participants to conditions.

consecutive behaviors or events, where good behavior 'A' leads to less desirable behavior 'B'. However, there is no reason to assume that the process of moral licensing actually operates in the order of these two consecutive behaviors. For instance, people may not consciously feel that after displaying certain good behavior 'A' they can now engage in undesirable behavior 'B', because of the prior good behavior. Perhaps people who face a dilemma in which they would like to engage in undesirable behavior 'B' (e.g., not donating money to charity) are more likely to find a reason why that is acceptable after having just performed a good action (e.g., performing voluntary work in a soup kitchen). In other words, moral licensing could also be a justification strategy that people deliberately use to excuse their morally questionable behaviors. This line of reasoning is illustrated by recent studies on moral credentialing. Merritt et al. (2012) found that people strategically attempt to earn moral credentials when they anticipate performing morally dubious behaviors. For instance, participants who expected that their future behavior could be regarded as prejudiced exaggerated how favorably they perceived a black person in a previous job hiring task. In a similar vein, Effron (2014) found that participants who were worried that their future behavior could be regarded as prejudiced or unethical overestimated to which extent previous nonracist choices or ethical behaviors proved their morality to other persons. Moreover, in their research on counterfactual licensing, Effron et al. (2012; 2013) showed that in order to justify future undesirable behaviors, people exaggerate negative counterfactuals of their foregone behavior.

Thus, instead of the reasoning being "I just did good deed A, so now I can do bad deed B", the reasoning might also be "I feel tempted to do bad deed B. Can I do that? Yes, because I just did good deed A". This might seem like a negligible nuance, but it also implies that not only aspects of the prior good deed (as many theories and moderators now focus on), but also tempting aspects of the morally questionable behaviors might be an important part of moral licensing theory. This reasoning fits with a justification-based account of self-regulation (De Witt-Huberts, Evers, & De Ridder, 2014a). Further studies on moral licensing might benefit from focusing on aspects of the immoral behavior as well.

Conclusion

Going back to the main goal of conducting the meta-analysis, the best estimate we have of the moral licensing effect is a Cohen's *d* effect size of 0.31. The effect is somewhat smaller than other typical effects in social psychology (Richard et al., 2003), but also relatively small effects can have large societal implications. An important consequence of our effect size estimate is that future studies on moral licensing need far more participants to allow for more solid conclusions. Especially when we want to delve further into the process of why this interesting phenomenon occurs, and what its boundary conditions are, the study power needs to be sufficient. The current meta-analysis gives researchers a good starting point to determine that power. The effects moral licensing has on less desirable and negative behaviors show the importance of continuing the research on this topic: A psychological process that helps to predict when people display ambiguously racist attitudes, engage in cheating behavior, and become more selfish, deserves further attention.

Appendix 2.1. Overview of non-eligible studies

Authors	Study	Reason for exclusion
Blanken, Van de Ven, and Zeelenberg (2012)	13	Evaluative judgment about behaviors is not a behavioral dependent variable.
Bradley-Geist, King, Skorinko, Hebl, and McKenna (2010)	3	Statistical information to calculate effect sizes was lacking.
Bradley-Geist, King, Skorinko, Hebl, and McKenna (2010)	4	The comparison between the two acquaintance conditions was not included because statistical information to calculate effect sizes was lacking.
Bradley-Geist, King, Skorinko, Hebl, and McKenna (2010)	5	Emotionality is not a behavioral dependent variable.
Brañas-Garza, Bucheli, Pas Espinosa and García- Muñoz (2013)	1	It was not possible to calculate effect sizes that are comparable to Cohen's *d* because the authors used a repeated measures within-subjects design.
Brewer, Cuite, Herrington, and Weinstein (2007)	1	Getting vaccinated is not necessarily a licensing induction. Instead, getting vaccinated may alter risk perception.
Cain, Loewenstein, and Moore (2010)	2	Perceived ethicality is not a behavioral dependent variable.
Chiou, Yang, and Wan (2011)	All	The studies are on self-regulation, not on moral behavior.
De Witt-Huberts, Evers, and De Ridder (2012)	1	The studies are on self-regulation, not on moral behavior. In addition, perceived effort is not a behavioral dependent variable.

Effron (2014)	All	Exaggerating previous good behavior when threatened is no behavioral dependent variable.
Effron, Monin, and Miller (2012)	1, 2	Evaluative judgments about other people's evaluations and perceptions about behavior are no behavioral dependent variables.
Effron, Monin, and Miller (2012)	4	Estimation of previous nonracist behavior is not a licensing dependent variable.
Effron, Monin, and Miller (2012)	5, 6	Perceived racism is not a behavioral dependent variable.
Effron, Monin, and Miller (2013)	All	The studies are on self-regulation, not on moral behavior.
Eskine (2012)	1	The studies are on self-regulation, not on moral behavior. In addition, an important precondition of licensing is that participants perceive their preceding behaviors as 'good'; this is not necessarily the case with exposure to a product.
Fishbach and Dhar (2005)	All	The studies are on self-regulation, not on moral behavior.
Gneezy, Imas, Brown, Nelson, and Norton (2012)	All	In the costless prosocial behavior condition, a donation was made on behalf of the participants. Thus, this condition did not include good behavior of the participants themselves. In the costly prosocial condition, a donation was deducted from the payment of their participants. Participants do not necessarily perceive this donation as 'good' but as being disadvantaged by the experimenter. Since the behaviors in both conditions are not necessarily interpreted as 'good' behaviors, we decided to exclude both conditions.
Jones-Doyle (unpublished)	1	The study is about observing licensing, not about self-licensing.
Jordan and Monin (2008)	1, 2	Indicating one's own morality is not a behavioral dependent variable.
Jordan, Mullen, and	1	Indicating one's own morality is not a behavioral dependent variable.

Murningham (2011)		
Khan and Dhar (2007)	All	The studies are on self-regulation, not on moral behavior.
Klotz and Bolino (2012)	All	A model on moral licensing related to citizenship and counterproductive work behavior is proposed, but not empirically tested.
Koritzky (2012)	1, 2	Benefitting the other player in the first game can be a strategy rather than licensing.
Kouchaki (2011)	All	Vicarious moral licensing is no self-licensing.
Landy (2013)	All	Evaluative estimations how much one is liked by other people is not a behavioral dependent variable.
Leonard (2012)	Conditions on food consumption	The conditions on food consumption are not included because these conditions are on self-regulation, not on moral behavior.
May and Irmak (2014)	All	The studies are on self-regulation, not on moral behavior. In addition, the studies measure the memories consulted by the participants when participants have the opportunity to indulge, rather than previous good behavior.
Merritt, Effron, Fein, Savitsky, Tuller, and Monin (2012)	All – except for Study 2 follow up data	The licensing dependent variable is not preceded by good behavior; instead, participants know in advance that they may have to display questionable behaviors and therefore they strategically obtain moral credentials.
Mukhopadhyay, Sengupta, and Ramanathan (2008)	All	The studies are on self-regulation, not on moral behavior.
Mukhopadhyay and Venkataramani (2009)	All	The studies are on self-regulation, not on moral behavior.
Ong, Mayer, and Tost (2014)	All	There is no self-licensing because participants do not perform good behaviors themselves.

Panzone, Wossink, and Southerton (2013)	All	From these correlational data, it was not possible to calculate an effect size that we could include in the meta-analysis.
Polman, Pettit, and Wiesenfeld (2013)	All	Self-licensing is not measured; this is about the judgment of other's transgressions.
Rynn and Vendello (2013)	All	An important precondition of licensing is that participants perceive their preceding behaviors as 'good'; this is not necessarily the case with a perceived underdog status.
Stewart (2012)	All	Perceived racism is not a behavioral dependent variable.
Ule, Schram, Riedl, and Cason (2009)	1	Cooperating often enough to keep a good reputation, but detecting after cooperation can be a strategic tactic rather than licensing.

References Appendix 2.1

Blanken, I., Van de Ven, N., & Zeelenberg, M. (2012). [*A variety of studies on the self-licensing effect*]. Unpublished raw data.

Bradley-Geist, J. C., King, E. B., Skorinko, J., Hebl, M. R., & McKenna, C. (2010). Moral credentialing by association: The importance of choice and relationship closeness. *Personality and Social Psychology Bulletin, 36*, 1564-1575. doi: 10.1177/0146167210385920

Branãs-Garza, P., Bucheli, M., Paz Espinosa, M., & García-Munoz, T. (2013). Moral cleansing and moral licenses: Experimental evidence. *Economics and Philosophy, 29,* 199-212. doi:10.1017/s0266267113000199

Brewer, N. T., Cuite, C. L., Herrington, J. E., & Weinstein, N. D. (2007). Risk Compensation and vaccination: Can getting vaccinated cause people to engage in risky behaviors? *Annals of Behavioral Medicine, 34,* 95-99. doi: 10.1007/bf02879925

Cain, D. M., Loewenstein, G., & Moore, D. A. (2010). When sunlight fails to disinfect: Understanding the perverse effects of disclosing conflicts of interest. *Journal of*

Consumer Research, 37, 836-857. doi: 10.1086/656252

Chiou, W. B., Yang, C. C., & Wan, C. S. (2011). Ironic effects of dietary supplementation: Illusory invulnerability created by taking dietary supplements licenses health-risk behaviors. *Psychological Science, 22*, 1081-1086. doi: 10.1177/0956797611416253

De Witt-Huberts, J. C., Evers, C., & De Ridder, D. T. D. (2012). License to sin: Self licensing as a mechanism underlying hedonic consumption. *European Journal of Social Psychology, 42,* 490-496. doi: 10.1111/j.1751-9004.2010.00263.x

Effron, D. A. (2014). Making mountains of morality from molehills of virtue: Threat causes people to overestimate their moral credentials. *Personality and Social Psychology Bulletin, 40,* 972-985. doi: 10.1177/0146167214533131

Effron, D. A., Monin, B., & Miller, D. T. (2012). Inventing racist roads not taken: The licensing effect of immoral counterfactual behaviors. *Journal of Personality and Social Psychology, 103*, 916-932. doi: 10.1037/a0030008

Effron, D. A., Monin, B., & Miller, D. T. (2013). The unhealthy road not taken: Licensing indulgence by exaggerating counterfactual sins. *Journal of Experimental Social Psychology, 49*, 573-578. doi: 10.1016/j.jesp.2012.08.012

Eskine, K. J. (2012). Wholesome foods and wholesome morals? Organic foods reduce prosocial behavior and harshen moral judgments. *Social Psychological and Personality Science, 4*, 251-254. doi: 10.1177/1948550612447114

Fishbach, A., & Dhar, R. (2005). Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *Journal of Consumer Research, 32*, 370-377. doi: 10.1086/497548

Gneezy, A., Imas, A., Brown, A., Nelson, L. D., & Norton, M. I. (2011). Paying to be nice: Consistency and costly prosocial behavior. *Management Science, 58*, 179 187. doi: 10.1287/mnsc.1110.1437

Jones-Doyle, J. (2012). *Moral licensing and environmentally friendly green behavior: Why people go green* (Master's thesis). Retrieved from http://spectrum.library.concordia.ca/974620/1/Jones-Doyle_MSc_F2012.pdf

Jordan, A. H., & Monin, B. (2008). From sucker to saint: Moralization in response to self threat. *Psychological Science, 19,* 809-815. doi: 10.1111/j.1467

9280.2008.02161.x

Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin, 37*, 701-713. doi: 10.1177/0146167211400208

Khan, U., & Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research, 43*, 259-266. doi: 10.1509/jmkr.43.2.259

Klotz, A. C., & Bolino, M. C. (2012). Citizenshp and counterproductive work behavior: A moral licensing view. *Academy of Management Review, 38*, 292-306. doi: 10.5465/amr.2011.0109

Koritzky (2012). Inequity aversion: A social utility or a reasonable strategy? Retrieved from http://www-bcf.usc.edu/~koritzky/InequityAversion_Koritzky.pdf

Landy, J. (2013). Unpublished raw data.

Kouchaki, M. (2011). Vicarious moral licensing: The influence of others' past moral actions on moral behavior. *Journal of Personality and Social Psychology, 101*, 702 715. Doi:10.1037/a0024552

Leonard, B. (2012). Unpublished raw data.

May, F., & Irmak, C. (2014). Licensing indulgence in the present by distorting memories of past behavior. *Journal of Consumer Research, 41,* 624-641. doi: 10.1086/676981

Merritt, A. C., Effron, D. A., Fein, S., Savitsky, K. K., Tuller, D. M., & Monin, B. (2012). The strategic pursuit of moral credentials. *Journal of Experimental Social Psychology, 48,* 774-777. doi: 10.1016/j.jesp.2011.12.017

Mukhopadhyay, A., Sengupta, J., & Ramanathan, S. (2008). Recalling past temptations: an information-processing perspective on the dynamics of self control. *Journal of Consumer Research, 35*, 586-599. doi: 10.1086/591105

Mukhopadhyay, A., & Johar, G. V. (1009). Indulgence as self-reward for prior shopping restraint: A justification-based mechanism. *Journal of Consumer Psychology, 19*, 334-345. doi: 10.1016/j.jcps.2009.02.016

Ong, M., Mayer, D. M., & Tost, L. P. (2014). Does working for a socially responsible company make employees more or less prosocial? The role of meaning in reconciling the moral licensing versus consistency debate. Retrieved from:

http://www.ivey.uwo.ca/cmsmedia/439295/Ong-manuscript.pdf

Panzone, L. A., Wossink, A., & Southerton, D. (2013). The design of an environmental index of sustainable food consumption: A pilot study using supermarket data. *Ecological Economics, 94*, 44-55. doi: 10.1016/j.ecolecon.2013.07.003

Polman, E., Pettit, N. C., & Wiesenfeld, B. M. (2013). Effects of wrongdoer status on moral licensing. *Journal of Experimental Social Psychology, 49*, 614-623. doi: 10.1016/j.jesp.2013.03.012

Rynn, D., & Vandello, J. A. (2013, January). Effects of underdog status on moral licensing. Poster session presented at the Society for Personality and Social Psychology, New Orleans.

Stewart, A. (2012). Unpublished raw data

Ule, A., Schram, A., Riedl, A., & Cason, T. N. (2009). Indirect punishment and generosity toward strangers. *Science, 326*, 1701-1704. doi: 10.1126/science.1178883

Appendix 2.2: Formulas used to calculate effect sizes

Cohen's *d* based on means and standard deviations was calculated as follows (Cohen, 1988);

$$d = \frac{\bar{X}1 - \bar{X}2}{s*},$$

where $s* = \sqrt{\frac{(n1-1)s_1^2 + (n2-1)s_2^2}{n1+n2-2}}$

In this equation, $\bar{X}1$ is the mean score of the self-licensing group, $\bar{X}2$ is the mean score of the control group and $s*$ is the pooled standard deviation, $n1$ is the sample size of the self-licensing group, $n2$ is the sample size of the control group, The pooled standard deviation includes the Hedges'g correction (Hedges, 1981). When specific information about the number of participants per cell was missing, we mailed the authors to ask for the exact numbers of participants per cell. If these numbers were still unknown, we divided the total amount of participants by the number of conditions.

~

Cohen's *d* based on *t* values was calculated as follows (Rosenthal & Rosnow, 1991);

$$d = \frac{t*2}{\sqrt{n-2}}$$

In this equation, *t* is the *t*-value of the *t*-test and *n* is the total sample size.

~

For Cohen's *d* based on means and standard deviations and Cohen's *d* based on *t* values, within-study variances (s^2) were calculated through the following formula (Hedges & Olkin, 1985);

$$s^2 = \frac{n1+n2}{n1n2} + \frac{d^2}{2(n1+n2)}$$

To convert effect sizes for studies wherein the dependent variable consisted of a binary choice (for instance, a choice between a virtuous and vicious option), the Phi coefficient ϕ was calculated (Cramer, 1946);

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

In this equation, χ^2 is the outcome of the chi square test (reported in the paper or calculated by hand), and n is the total sample size. Next, we converted Φ to Cohen's d through the following formula (Rosenthal, 1994);

$$d_{\phi} = \frac{\phi}{\sqrt{1-\phi^2}} \sqrt{\frac{df\,(n1+n2)}{n1n2}}$$

In this equation, $df = n1 + n2 - 2$.

~

For Cohen's d based on the Φ coefficient, within-study variances (s^2) were calculated as follows (Stuart & Ord, 1994);

$$s^2_{d\phi} = \frac{n1+n2}{n1n2\left(1-\phi^2\right)^2}$$

The 95% confidence intervals of the Cohen's d effect size were computed in the following way (Viechtbauer, 2007);

$$\mathrm{CI}[95\%] = d \pm 1.96 \sqrt{\frac{n1+n2}{n_1 n_2} + \frac{d^2}{2m}}$$

We applied the small sample bias correction (Lipsey and Wilson, 2004) for all effect sizes;

$$d'_{sm} = d_{sm}\left[1 - \frac{3}{4N - 9}\right]$$

References Appendix 2.2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cramer, H. (1946). *Mathematical methods of statistics*. Princeton: Princeton University Press.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128. doi:10.3102/10769986006002107.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Thousand Oaks, CA: Sage.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York: Russel Sage Foundation.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed ed.). New York: McGraw Hill.

Stuart, A., & Ord, K. (1994). *Kendall's advanced theory of statistics. Vol. I: Distribution theory (6th ed.)*. New York: Edward Arnold.

Viechtbauer, W. (2007). Approximate confidence intervals for standardized effect sizes in the two-independent and two-dependent samples design. *Journal of Educational and Behavioral Statistics, 32,* 39-60. doi:10.3102/1076998606298034

CHAPTER 3

A *p*-Curve Analysis of Moral Licensing

In this chapter we present a *p*-curve analysis that examines the evidential value of the moral licensing effect. The results show that the current body of research on moral licensing does not have evidential value; based on the distribution of *p*-values the published moral licensing tests do not provide support for the hypothesis that previous moral behavior makes it more likely that one subsequently engages in immoral, unethical, or otherwise problematic behavior. Since the *p*-curve analysis shows no evidential value for the moral licensing effect in the current body of empirical research, we strongly recommend researchers studying moral licensing to 1) use larger samples to study the effect, 2) first establish a reliable paradigm to study licensing effects before exploring moderators, and 3) start with revising the theoretical framework of moral licensing so it becomes studied less as an effect, but more as a theory.

This chapter is based on: Blanken, I., Van Wolferen, J., Van de Ven, N., & Zeelenberg, M. (2015). A *p*-curve analysis of moral licensing. *Manuscript under review.*

Moral licensing is defined as the phenomenon that when people behave in a moral way, they are later more likely to display behaviors that are immoral, unethical, or otherwise problematic (e.g., Merritt, Effron, & Monin, 2010). For instance, participants who bought ecofriendly products in a virtual shopping paradigm subsequently offered less money to other people in an ultimatum game and stole more money compared to participants who bought regular products (Mazar & Zhong, 2010). Moral licensing is related to behaviors in various domains such as racism (Choi & Crandall, & La, 2014; Effron, Cameron, & Monin, 2009; Effron, Monin, & Miller, 2012; Mann, & Kawakami, 2012), job hiring (Cascio & Plant, 2015; Monin & Miller, 2001), ethics (Conway & Peetz, 2012; Jordan, Mullen, & Murnighan, 2011), and consumerism (Khan & Dhar, 2006). As moral licensing has adverse consequences for such a wide range of behaviors, it is an important phenomenon to study. However, recently some doubt arose concerning the robustness of this phenomenon (see Chapter 2 and 3). In the current chapter we conduct a *p*-curve analysis (Simonsohn, Nelson, & Simmons, 2014a,b) to test the strength of the evidential value of the current body of empirical studies on moral licensing.

The robustness of the moral licensing effect was called into question most clearly by the meta-analysis from Chapter 2. This analysis included 91 experimental studies, in which initial moral or good behavior was manipulated and the effect of this manipulation on immoral or decreased moral behavior was measured. The average effect size of moral licensing was found to be small-to-medium (Cohen's d = 0.31). Furthermore, this analysis did not find any statistical support for moderators that were theorized to be of importance. For instance, this meta-analysis did not show that licensing conditions (e.g., recalling moral behavior) contrasted with negative control conditions (e.g., recalling immoral behavior) led to larger effects than licensing conditions contrasted with neutral control conditions (e.g., recalling neutral behavior, such as doing groceries). Theoretically, one would expect a larger difference between a licensing condition and a negative control condition. The absence of such effect was surprising, especially in the light of ideas on the moral cleansing effect in which recalling previous immoral behavior leads to more subsequent moral behavior (the

exact opposite of the moral licensing effect; see Conway & Peetz, 2012; Jordan et al., 2011; Zhong & Liljenquist, 2006).

Strikingly, the only moderator that was confirmed in the meta-analysis was that published studies had larger effects (average d = 0.43) than did unpublished studies (average d = 0.11). This finding was in line with the regression test for funnel plot asymmetry that indicated the presence of a positive publication bias. In the current chapter, we continue this investigation of the robustness of the moral licensing effect, by submitting the published findings to a *p*-curve analysis. In short, a meta-analysis is good at testing for moderators, but not good at correcting for publication bias and typically leads to overestimations of the true effect size. A *p*-curve cannot test for moderators, but is not influenced by publication bias (Simonsohn et al., 2014a).

Why use *p*-curve?

Creating a *p*-curve is a meta-analytic technique that tests the evidential value of a set of published studies by analyzing the shape of the distribution of their *p*-values that are < .05 (Simonsohn et al., 2014a). Unlike traditional meta-analytic techniques, *p*-curve is able to provide an effect size estimate that is not influenced by publication bias (Simonsohn et al., 2014b). Even when procedures to correct for publication bias are applied, traditional meta-analytic techniques are biased towards the conclusion that an effect exists. *P*-curve provides a publication-bias-free answer to the question of whether an effect exists. If *p*-curve concludes a set of studies has evidential value, it can also produce a publication-bias-free estimate of the size of the effect. Logically, when *p*-curve concludes a set of studies does not have evidential value, the estimate for the effect size is zero. We refer interested readers to the original *p*-curve articles of Simonsohn et al. (2014a,b) for information about the specifics of the analysis and restrict ourselves to a brief explanation of how *p*-curve tests for evidential value.

How does *p*-curve test for evidential value?

The *p*-curve tests the evidential value of a set of studies: it tests whether a set of findings reflects a *true* effect (effect-size > 0). It does so based on the distribution of *p*-values between 0 and .05. It allows researchers to determine whether there are many

'small' (e.g., < .01) *p*-values, whether the *p*-values are approximately equally distributed across the range that is generally considered to be 'significant' (< .05), or whether there are many 'large' (e.g., between .04 and .05) *p*-values. Each of the three outcomes described above allows researchers to draw a particular conclusion about the *p*-curved set of findings. We will discuss these three possible outcomes below.

First, if there are many small *p*-values, the conclusion is that the set of *p*-curved findings has evidential value. The most intuitive way to understand why this is the case is to use the following example. Think about an effect we know to be true: In the United States, on average, men are taller than women.[8] If one were to run 100 studies (with sufficient power) comparing the height of men and women, it is likely that one will find a 'very significant' effect in most studies. There would not be many studies that are not significant, and only a few would produce *p*-values just below .05. Figure 3.1 shows that plotting this distribution leads to a right-skewed distribution (the *p*-curve is downward sloped) that looks like the dashed line; small *p*-values are more likely to be obtained than larger ones. Therefore, when *p*-curve detects a distribution of *p*-values that is right-skewed, it will conclude that the set of studies that produced the *p*-values has evidential value.

Second, the distribution of *p*-values can be uniform (the *p*-curve is horizontal): approximately equal amounts of small and large *p*-values. The conclusion that can be drawn from such a set of *p*-values is that the combined set of studies does not have evidential value. In other words, the set of studies does not provide support for the hypothesis that the effect exists. The intuition lies in the definition of the *p*-value. *P*-values indicate the likelihood that one would observe an effect at least as extreme as the one observed, *if there were truly no effect*. Going back to the example of running 100 studies, if they test a non-existing effect $p < .05$ would be observed in 5% of the studies, and $p < .04$ in 4% of the studies. Logically, $.04 < p < .05$ would occur 1% of the studies. Similarly, $.01 < p < .02$ would also occur in 1% of the studies. Put differently, when there really is no effect, each *p*-value is equally likely to be observed. Therefore, if a *p*-curve

[8] This example is also used in the original *p*-curve paper (Simonsohn et al., 2014a) and we copy it here because it works so well.

detects a uniform distribution of *p*-values, it concludes that the studies do not have evidential value as the results are likely to have existed due to chance. The solid line in Figure 3.1 represents this outcome.

Third, if there are more large *p*-values than there are small ones, the conclusion is that the set of *p*-curved findings has no evidential value, and is likely to be *p*-hacked. *P*-hacking refers to the selective reporting of analyses that produce significance ($p < .05$). It is often possible to 'detect' a significant effect even where there truly is no effect (Simmons, Nelson, & Simonsohn, 2011), for example by selectively dropping observations or by selecting the measures or studies that are being reported. The intuition for why this leads to an abundance of *p*-values just below .05 is that researchers search for significant effects in their data until they find one. Any *p*-value below .05 is theoretically publishable, so researchers are unlikely to keep searching until they find a very small *p*-value (e.g., < .01) when $p = .049$ 'does the job'.[9] This way of analyzing data leads researchers to falsely conclude that there is an effect, and it results in a left-skewed distribution that looks like the dotted line in Figure 3.1. Therefore, when *p*-curve detects a distribution of *p*-values that is left-skewed (the *p*-curve is upward sloped), it will conclude that the set of studies that produced the *p*-values has no evidential value and is likely *p*-hacked.

[9] This is why *p*-curve analyses only include *p*-values below .05. Greater *p*-values could technically be used to test for evidential value. However, to determine whether a set of studies has been *p*-hacked *p*-curve assume that researchers *p*-hack 'until they get $p < .05$'. Therefore, $p = .05$ serves as the upper boundary of *p*-values that can be included in *p*-curve.

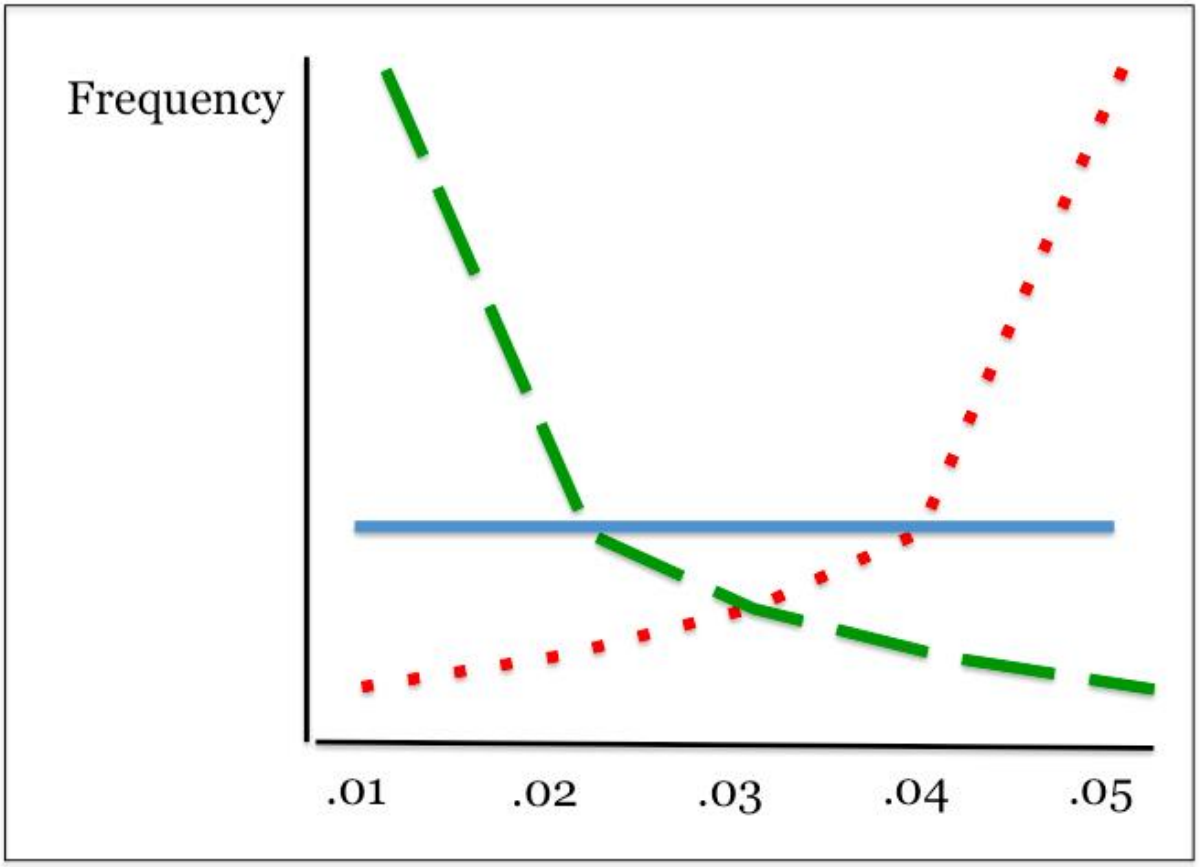


Figure 3.1. Three possible distributions of significant p-values. The *y*-axis displays the frequency of *p*-values and the *x*-axis displays the value of the observed *p*-values. *P*-curve tests the evidential value of a set of studies by determining whether the shape of the distribution of significant *p*-values looks like one of the shapes above. Sets of studies that have evidential value (i.e., that test true effects) produce distributions that look like the right-skewed dashed line. Sets of studies that do not have evidential value, but are not *p*-hacked produce uniform distributions like the flat solid line. Sets of studies that do not have evidential value and are *p*-hacked produce distributions like the left-skewed dotted line.

The *p*-curve analysis produces a significance test for each of these three possible outcomes. These tests answer the following three questions: (1) Is the observed distribution significantly more right-skewed than a uniform (flat) distribution? (2) Is the observed distribution significantly flatter (or more left-skewed) than if the *p*-curved studies had 33% power? (3) Is the observed distribution significantly more left-skewed than a uniform (flat) distribution? Test number two requires a bit more explanation: The shape of *p*-curve depends on the statistical power of the studies that are submitted to it. The *p*-curve becomes more right-skewed as the power increases.

One intuitive way of thinking about this feature of *p*-curve is to go back to the example of studies that test whether men are taller than women. Imagine you run 100 studies with 20 men and 20 women (not so much power), and 100 studies with 1.000 men and 1.000 women (a lot of power). The set of studies with N = 2.000 will produce many more small *p*-values than the set of studies with N = 40. The distribution of the N = 2.000 set of studies would thus be much more right-skewed than that of the N = 40 set.

Interestingly, for a given level of power, *p*-curve is almost the same for every underlying sample-size and effect-size combination (Simonsohn et al., 2014a). Therefore, *p*-curve 'knows' what the distribution of significant *p*-values in a set of studies would be if they had 33% power: it would be just slightly right-skewed. The second statistical test compares whether the observed distribution of *p*-values is significantly flatter than a line of a true effect being studied with 33% power would be. If the test indicates that the line is flatter than a line with 33% power would look like, the *p*-curve analysis concludes that the set of studies does not have evidential value. The 33% power cut-off is arbitrary (just as $p < .05$ to determine significance is arbitrary), but it is sensible: If the distribution of significant *p*-values is flatter than one that would be produced by studies that would only have a 1 in 3 chance to find an effect (if it exists), it makes sense to conclude that the set of studies do not provide evidence for the claims they test. It is important to note that the conclusion if this test is significant is *not* that the effect the studies test is false. The conclusion is merely that the complete set of studies that forms the *p*-curve does not provide evidence for the effect, as it is likely to have arisen based on chance.

Finally, because the shape of the *p*-curve is so closely related to the power of the studies that are submitted to it, *p*-curve also gives an estimate of the average power of that set of studies. It is worth noting that *p*-curve itself always has more power than the studies that are submitted to it. With 20 *p*-values, it has more than 60% power to conclude the studies have no evidential value. With 20 *p*-values, it also has more than 80% power to conclude the studies have evidential value, even if those studies themselves have 33% power. In addition, its false positive rate is 5% (because we use $\alpha = .05$ as the criterion for significance), but the actual false positive rate diminishes as the effect size in the

set of studies submitted to *p*-curve rises (Simonsohn et al., 2014a). Put differently, when there is no effect the false positive rate is at its highest (5%). This implies that the test that tests for evidential value is powerful, while the test that tests whether the set of studies does not have evidential value is very conservative. This makes it hard for *p*-curve to get it wrong.

To summarize, we present a *p*-curve analysis that examines the evidential value of the combined set of studies documenting the moral licensing effect. *P*-curve provides an alternative method to investigate whether the published studies testing for moral licensing included in the meta-analysis in Chapter 2 reflect a true effect. This is important because in Chapter 2 we noted that there is a publication bias, something a meta-analysis can not fully correct for. A *p*-curve is not affected by publication bias and can thus provide a better effect size estimate in those cases (Simonsohn et al., 2014a). If *p*-curve concludes that the studies that test for moral licensing have evidential value, we will also estimate the effect size using *p*-curve.

Methods and results

Inclusion Criteria. We used the dataset from the meta-analysis on moral licensing from Chapter 2. In this meta-analysis, the literature search was based on the definition of Merritt et al. (2010, p. 344): "Past good deeds (or good intentions) liberate individuals to engage in behaviors that are immoral, unethical, or otherwise problematic, behaviors that they would otherwise avoid for fear of feeling or appearing less moral". Studies were included if 1) the behavior that was measured met this definition of moral licensing and 2) reported statistics were adequate to calculate effect sizes. For a more detailed description of the inclusion criteria, see Chapter 2.

For the current *p*-curve analysis, *p*-values were included if they met the following criteria: 1) The *p*-value is published in a scientific journal[10]; 2) The *p*-value reflects a

[10] One could argue that studies that passed peer-review might be better, more robust, or otherwise superior to studies that have not passed peer-review. In that sense, the current test of the moral licensing effect is more conservative than the meta-analysis in Chapter 2, which also included unpublished studies.

crucial moral licensing test; 3) The *p*-value is significant. The second criterion implies that only *p*-values that reflect a comparison between a moral licensing and a control condition are included. *P*-values reflecting a comparison between more than these two conditions or interactions with other variables are therefore not included.

It is important to note that the second criterion also has implications for the conclusions we can and cannot draw. The articles we *p*-curved sometimes make predictions about interactions, or about things that are unrelated to licensing. For example, Conway and Peetz (2012) expected more prosocial intentions if participants recalled distant moral behavior than distant immoral behavior (a consistency effect), and more prosocial intentions if participants recalled recent immoral behavior than recent moral behavior (a compensatory effect). The crucial licensing test would be to compare participants who recalled recent moral behavior with participants who recalled recent immoral behavior. However, the results of this specific test are not reported. Rather, a contrast analysis was reported to test the interaction in which the moral-distant conditions and immoral-recent condition would lead to greater prosocial intentions than the immoral-distant and moral-recent conditions. We do not *p*-curve those predictions, so our analysis does *not* test whether the studies contain evidential value for those predictions. The selection criteria ensure that we *only* test the evidential value of the empirical support for the basic effect of moral licensing.

In the process of including the *p*-values, we ran into a number of unexpected problems. Out of 54 published comparisons between a licensing and a control condition, 18 comparisons (33%) were not significant ($p > .05$). For 10 comparisons, the crucial licensing test was not reported (marked with 'x' in the Disclosure Table). For example, planned contrast tests to compare the moral licensing condition with two other conditions were reported rather than the test between the moral licensing and the control condition (Cascio & Plant, 2015). In four cases, recalculation of the statistical tests shows that tests that were reported as significant were actually not significant ($p > 0.05$, marked with 'xx' in the Disclosure Table). Furthermore, six comparisons included statistical oddities, such as reporting different *M*s and *SD*s for the same variable and larger *t*-values than those recalculated from the reported *M*s and *SD*s

(marked with 'xxx' in the Disclosure Table). This is consistent with earlier findings which show a high prevalence of statistical reporting errors in psychological science (Veldkamp, Nuijten, Dominguez-Alvarez, Van Assen & Wicherts, 2014).

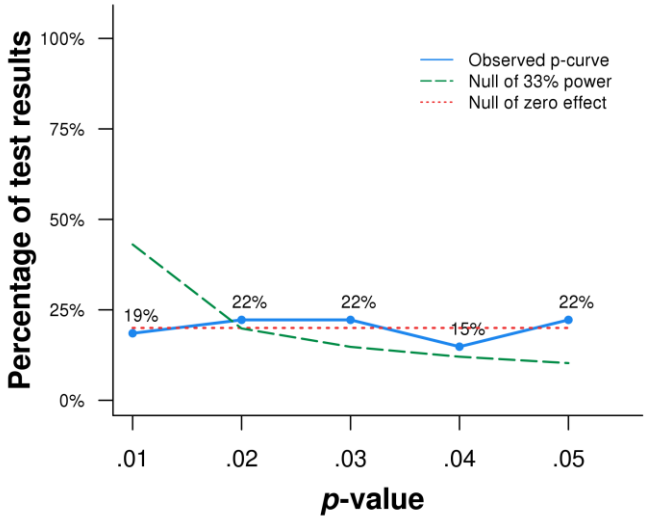
The final *p*-curve dataset consisted of 27 *p*-values from 27 studies reported in 13 articles, which in total were cited 1594 times.[11] An overview of the included *p*-values can be found in the Disclosure Table. All analyses were conducted using the online *p*-curve app 3.0 (http://w ww.p-curve.com /app3/).

Robustness check

The *p*-values that are submitted to *p*-curve should be independent of each other, so if a study has two dependent variables, only (the first) one can be submitted to the *p*-curve. Therefore, we report a second *p*-curve in which the *p*-value from the second dependent variable replaces the first. This was the case for 4 out of the 27 observations. This serves as a robustness-check and provides an indication of the stability of the results of the *p*-curve. The results of the original *p*-curve and the robustness check are reported in Table 3.1 and Figure 3.2.

[11] Google Scholar. June 2, 2015

P-curve



Robustness-check

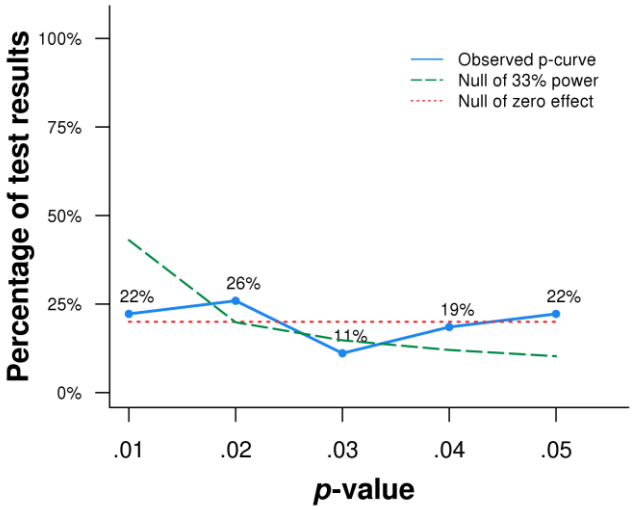


Figure 3.2. The distribution of *p*-values in moral licensing studies (solid line).

Table 3.1

Statistical results of p-curve and robustness check

	P-curve		***Robustness-check***	
	Binomial Test *(Share of significant results p <.025)*	**Continuous Test** *(Aggregate pp-values via Stouffer Method)*	**Binomial Test** *(Share of significant results p <.025)*	**Continuous Test** *(Aggregate pp-values via Stouffer Method)*
1) Studies contain evidential value. *(Right skew)*	*p* = .6494	Z = -0.75, *p* = .2263	*p* = .5000	Z = -2.02, *p* = .0219
2) Studies' evidential value, if any, is inadequate. *(Flatter than 33% power)*	*p* =.0114	Z = -2.18, *p* = .0145	*p* =.0296	Z = -0.93, *p* = .1774
3) Studies exhibit evidence of intense *p*-hacking. *(Left skew)*	*p* =.5000	Z = 0.75, *p* = .7737	*p* =.6494	Z = 2.02, *p* = .9781
Average power of tests included in *p*-curve *(correcting for publication bias)*	<5%		8%	

Note. The *p*-curve app provides two related tests: A binomial test and a continuous test. The binomial test simply tests whether the proportion of small (<.025) *p*-values is significantly (1) larger than 50%, (2) smaller than 50%, or (3) smaller than the percent that would be expected if the studies had 33% power. The continuous tests are based on Stouffer's Method we refer interested readers to Simonsohn et al. (2014a) for details. Put simply, this test analyzes whether the shape of the distribution is significantly (1) more right-skewed than a uniform distribution, (2) significantly more left-skewed than a mildly right-skewed distribution, or (3) more left-skewed than a uniform distribution.

Summary of results

Both the binomial and the continuous *p*-curve tests indicate that the studies' evidential value is inadequate (flatter than 33% power, see the dashed line in Figure 3.1.). There

is no indication of *p*-hacking (i.e., the distribution is not left-skewed). The results of the binomial tests from the robustness check confirm these results. The continuous test from the robustness check seems to indicate that the studies have some evidential value. However, even this test suggests minimal evidential value (average power = 8%) and qualitatively, both *p*-curves do not look like ones that a healthy literature would produce.

Figures 3.3 and 3.4 provide an indication of the robustness of the results. They show what the *p*-value of each test in Table 3.1 would be if the highest or lowest *k* original *p*-values would be deleted. These figures thus give an indication of the strength of the results and provide insight into whether particular results hinge on just one *p*-value. It shows that in the first *p*-curve, dropping the 4 highest original *p*-values would lead to significant right skew, and dropping the two highest original *p*-values would render the 33% test non-significant. In the robustness-check *p*-curve, dropping the lowest original *p*-value (so just one) would render the test for right skew non-significant, and dropping (that same) lowest original *p*-value would lead to a significant 33%-test.

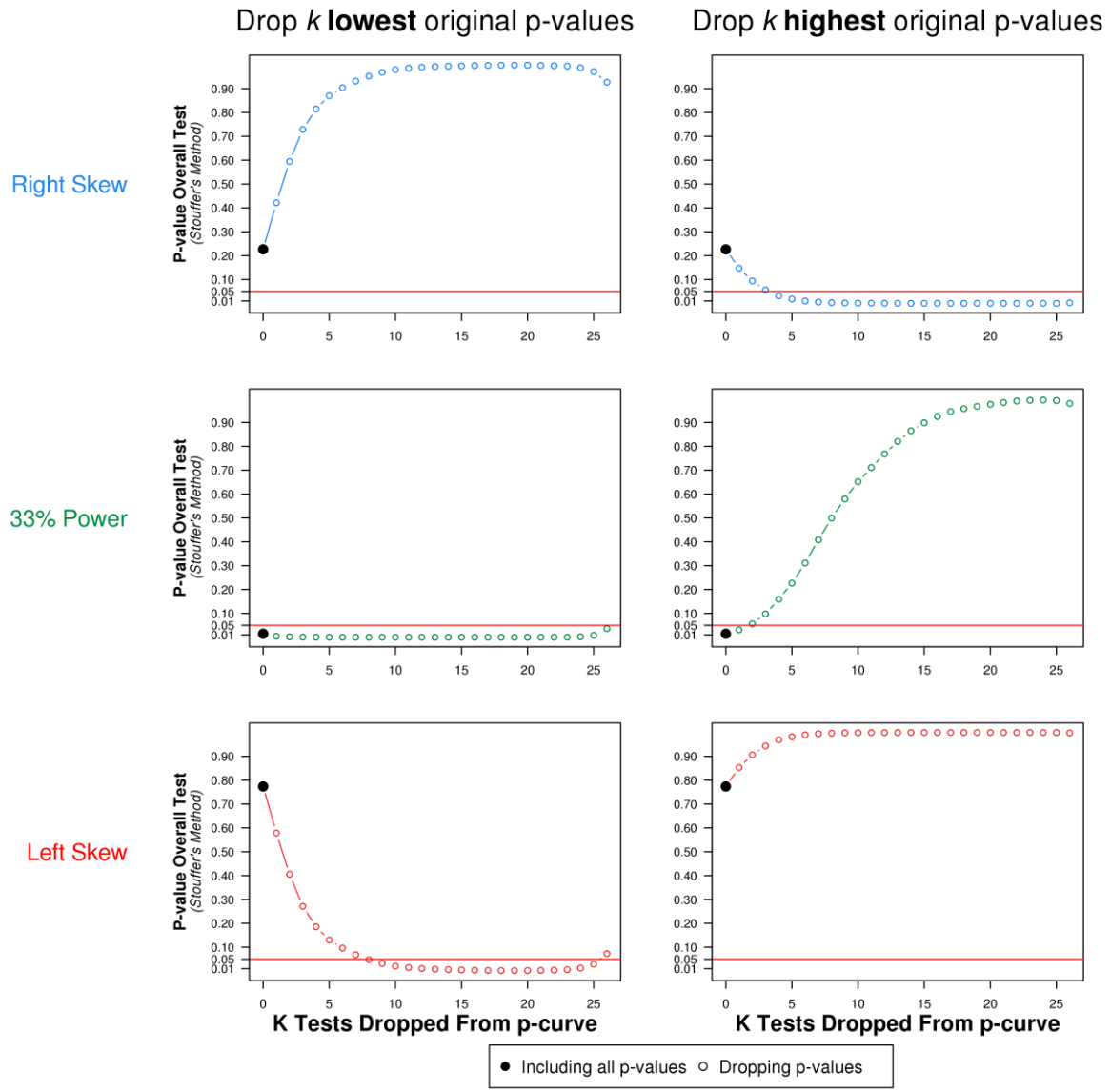


Figure 3.3. The effects of dropping k lowest or highest original *p*-values on the statistical tests in the *p*-curve.

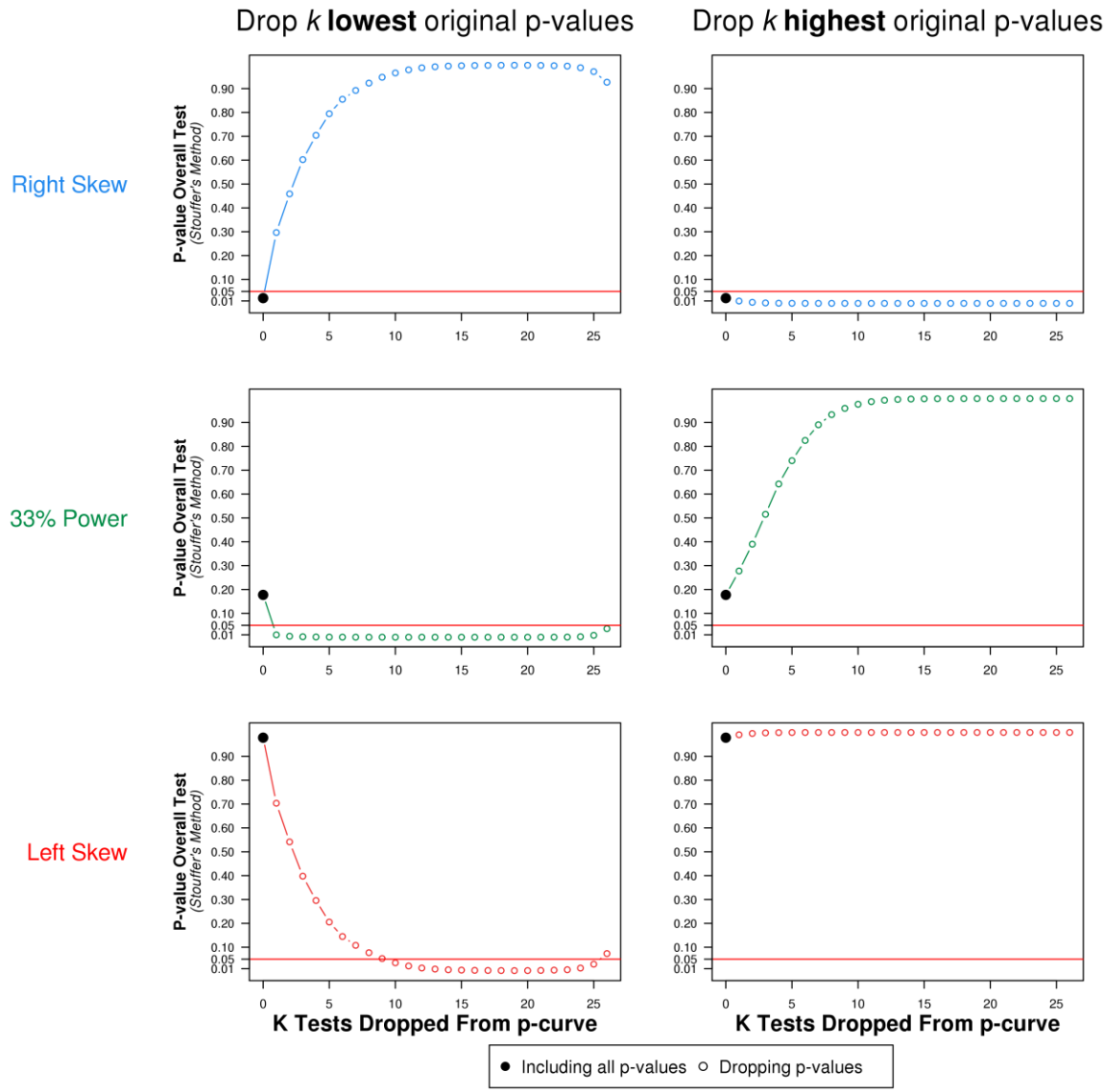


Figure 3.4. The effects of dropping k lowest or highest original *p*-values on the statistical tests in the robustness check *p*-curve.

Overall, we think the conclusion that the set of studies does not have evidential value is justified because 1) the initial *p*-curve concludes that the studies evidential value is inadequate, 2) the robustness check concludes that there is only 'minimal' evidential value in set of studies (average power = 8%) and this test hinges on just one *p*-value, and 3) we found many errors and misreporting in the studies. Below, we elaborate on the implications this conclusion has for the study of moral licensing.

Discussion

In the current chapter we conducted a *p*-curve analysis (Simonsohn et al., 2014a,b) to evaluate the evidential value of the published moral licensing tests reported in the meta-analysis in Chapter 2. We found that research on moral licensing lacks evidential value; the distribution of *p*-values shows that the complete set of published studies on moral licensing does not provide support for the hypothesis that previous moral behavior leads to immoral, unethical, or otherwise problematic behavior.

P-curve versus meta-analysis

The results from the *p*-curve analysis seem inconsistent with the meta-analysis on moral licensing (Chapter 2), which suggested that there is a small significant licensing effect (d = 0.31). A meta-analysis is a traditional approach to estimate the size of an effect and to test its moderators, but is not good at correcting for publication bias (Simonsohn et al., 2014b). The *p*-curve is a new method to provide an unbiased estimate of an effect, but since it has been introduced very recently it might have some as of yet unknown limitations. As outlined in the introduction, the meta-analysis on moral licensing revealed a positive publication bias, implying that more positive findings than null- or – negative- findings on moral licensing were published. This could have led to an overestimation of the effect size as it becomes much more likely that one missed some non-significant findings as they are more difficult to track down. The *p*-curve analysis is not influenced by such a publication bias, which could explain the different results between both methods. The more traditional meta-analyses are needed in addition to a *p*-curve analysis in order to draw a complete picture of an effect. *P*-curve analysis is not able to analyze moderators. Both methods thus complement each other.

Implications

The finding that current studies on moral licensing do not provide empirical support for the hypothesis that previous moral behavior leads to behavior that is less moral has large practical and theoretical implications for future research on moral licensing. It is important to note that we found no evidential value for the moral licensing effect *as it*

is currently being investigated. This does not necessarily imply that moral licensing does not exist: The *p*-curve cannot evaluate whether the theory of licensing is correct or incorrect, but only whether the body of empirical support for it jointly contains support for the existence of the effect. Therefore, we stress that the current analysis does not invalidate licensing theory. However, if parts of the theory are inspired by findings that are included in the *p*-curved set of findings, our *p*-curve analysis does cast doubt about the truth-value of the theory. We therefore strongly recommend researchers studying moral licensing to start afresh with investigating this phenomenon.

When starting afresh with investigating moral licensing, we think that it is vital to come up with a theory of licensing that integrates the licensing process with theories and findings that show the importance of consistency in behavior (Abelson et al., 1986; Gawronski & Strack, 2012), as they predict the opposite of a moral licensing effect (e.g., Conway and Peetz, 2012). Consistency theory predicts that people who perform a good deed are expected to engage in future good behavior, as people want to appear consistent. A prior good deed makes one see oneself as a good person, after which one becomes more likely to engage in future good behavior (Bem, 1972). Since good behavior is not always followed by less desirable behavior, researchers should clearly outline under which circumstances they predict licensing versus consistency effects before running experiments.

In addition, we strongly recommend researchers to conduct new studies on moral licensing with properly powered samples to allow for data with evidential value, as the *p*-curve analysis shows that the average power of the tests included in the *p*-curve is well below 10%. Studies on moral licensing need larger sample sizes to be able to reliably test for the effect and possible moderators. We think it would be helpful if researchers first establish the licensing effect reliably in a paradigm, before digging deeper into the phenomenon.

When developing a reliable paradigm for the moral licensing effect, we think it is very useful to adopt a full cycle approach (Cialdini, 1980) of self-licensing. Mortensen and

Cialdini (2010, p.53) argue that when studying a novel phenomenon researchers should first

> use naturalistic observation to determine an effect's presence in the real world, theory to determine what processes underlie the effect, experimentation to verify the effect and its underlying processes, and a return to the natural environment to corroborate the experimental findings.

We think the initial step, testing whether and how people experience licensing in their daily lives, has not extensively been made yet in the self-licensing literature. If we need to rethink the paradigms we use to study licensing, starting with examining how it exists in real life is a valuable first step.

Initial efforts have been made to investigate licensing effects outside of the traditional experimental paradigm and test for licensing effects in real life. Hofmann, Wineski, Brandt, and Skitka (2014), investigated the occurrence of moral and immoral behaviors in daily life. Consistent with theorizing on moral licensing, they found that individuals who commited a moral act earlier in a day were more likely to subsequently display immoral behavior, and less likely to subsequently display moral behavior. In addition, in Chapter 5 we find that people can recognize instances of licensing in their own past behavior. These past instances can be seen as fitting *good deed self-licensing*, where a good deed permits one to display undesirable behavior, or *temptation-based self-licensing*, where the temptation to display undesirable behavior initiates a search for a license. Recently, research on licensing in the self-regulatory domain found additional evidence for temptation-based self-licensing (Chapter 6; De Witt-Huberts, Evers, & De Ridder, 2014b). Since past research on moral licensing focused on manipulating good behavior and did not give sufficient attention to the role of the temptingness of displaying the undesirable behavior, we recommend future research to consider the option that moral licensing can also be initiated by the temptingness of the undesirable behavior.

Temptation-based self-licensing thus implies that not the prior good behavior, but the temptation to display undesirable behavior can trigger the process of self-licensing. This account of licensing seems more in line with theories on consistency, since people

who experience temptation-based licensing try to make their beliefs consistent with their desired behaviors (i.e., giving in to temptation) through searching for justifications. Specifically, temptation-based licensing shows some close resemblances with cognitive dissonance reduction, which implies that individuals want to reduce the negative feelings that they experience when they realize that their behaviors contradict their beliefs (Festinger, 1957). For observers, temptation-based licensing may seem to counteract consistency, but for actors it feels like making their attitudes and beliefs consistent with their desired behaviors (i.e., giving in to temptation).

Finally, we believe that researchers should reflect on established theories from different fields that can improve the understanding of moral licensing. Several theories on negative spillovers, where a second action offsets the effectiveness of a prior action or state have previously been established in the field of economics. Examples are the rebound effect (Greening, Greene, & Difiglio, 2000), the Jevons paradox (Jevons, 1865), moral hazard (Arrow, 1963), and risk homeostasis theory (Wilde, 1998). For instance, research on the rebound effect shows that when households receive a high-efficiency washing machine, they tend to increase the frequency of clothes washing, thereby offsetting the beneficial effects for the environment (Davis, 2008). This may closely reflect what happens when people feel 'licensed' to display undesirable behavior. We think that researchers in the field of licensing should therefore consider what these seemingly related theories can explain and predict about the nature of moral licensing.

Conclusion

The *p*-curve analysis in the current chapter finds that the existing body of research on moral licensing does not have evidential value to support the existence of moral licensing. Importantly, this does not necessarily imply that moral licensing does not exist, but it does imply that research has not been able to reliably document it as the findings are likely to have arisen due to chance. Researchers studying moral licensing are strongly recommended to start afresh with investigating this phenomenon by establishing a robust paradigm. Recent research shows that moral licensing effects occur in daily life (Chapter 5; Hofmann et al., 2014), but the process may work differently than previously assumed (Chapter 6; De Witt-Huberts et al., 2014b). More

research needs to establish the evidential value of this novel approach. We also think that when doing so, moral licensing theory should be integrated with related theories from different fields. The study of moral licensing should thereby hopefully shift from a focus on the moral licensing effect, to a theory of moral licensing that fits with other established theories.

Disclosure Table.

Author	**Quoted text from original paper indicating prediction of interest to researchers**	**Study design**	**Key statistical Result**	**Quoted text from original paper with statistical results**	**Result**	**Robustness result**	**Comments**
Bradley-Geist et al., 2010 Study 1 xxx	We predict that people who acquire moral credentials by choosing to cite positive experiences with a Black individual subsequently will be more likely than other people to show preferences for White over Black applicants in a job selection scenario.	Three cells, two control conditions	Difference of means	Supporting expectations, free-choice participants who were able to choose to write about a positive experience with a Black individual later expressed more bias toward White applicants (M = 4.29, SD = .64) than did those in the control condition (M = 3.86, SD = 0.57), t = -2.32, p < .05, and those in the forced-choice condition who were instructed to write about a positive experience (M = 3.76, SD = 0.72), t = -2.25, p <.05.	$t(34)$ = -2.32	$t(36)$ = -2.25	The reported t-values are larger than those recalculated* from the reported M and SDs. We included the p-values based on the reported t-values. Degrees of freedom were not reported so we calculated them ($n1$+$n2$-2).
Bradley-Geist et al., 2010 Study 2 x	Given that people feel more comfortable stereotyping Hispanics, we predicted that choice may not be as important of a factor in establishing credentials as it was for Study 1 when the target was Black.	Three cells, two control conditions	Difference of means	A MANOVA was first conducted on the job suitability and tolerance for prejudice variables with experimental condition as the independent variable. Wilks's Lambda was marginally significant (.98, p = .11), and Roy's largest root was statistically significant (.12, p < .05), so separate one-way ANOVAs were run on each of the dependent variables. Experimental condition was not significantly related to job suitability ratings, (2, 61) = 0.45, p > .10, but the pattern of means was consistent with the evaluation items such that participants in the control condition tended to express less prejudice (M = 4.05, SD = 0.67) than participants in the free-choice condition (M = 4.15, SD = 0.63). Experimental condition was significantly related to the tolerance for prejudice scale, F(2, 62) = 3.17, p			Prediction sounds like attenuated interaction, but there is no direct comparison with Black person conditions. The prediction could be interpreted as 'choice does not play a role', so both forced and free choice conditions would lead to a licensing effect (i.e., more prejudice towards Hispanics). However, the authors do not test the comparison between forced

				< .05. *t* tests revealed that individuals in the free-choice condition expressed greater tolerance for prejudice (*M* = 3.44, *SD* = 0.84) than did individuals in the control condition (*M* = 2.76, *SD* = 0.83), *t* = –2.76, *p* < .01, but the ratings did not differ significantly from those individuals in the forced-choice writing assignment condition (*M* = 2.90, *SD* = 0.84)			choice and the control condition. So it remains unclear what the authors predict. When calculating* the *p*-values for both DVs, we found that for the job hiring task DV, the comparison between the free choice condition (*M* = 4.15, *SD* = 0.63) and the control condition (*M* = 4.05, *SD* = 0.67) was not significant, *t*(40) = 0.50, *p* = .622. The comparison between the forced choice condition could not be calculated because there is no information about the mean of the forced choice condition. For the tolerance for prejudice scale DV, we found that the comparison between the free choice condition (*M* = 3.44, *SD* = 0.84) and the control condition (*M* = 2.76, *SD* = 0.83) was significant, (40) = 2.64, *p* = .012. The comparison between the forced choice condition (*M* = 2.90, *SD* = 0.84) and the control condition was not significant, *t*(42) = 0.56, *p* = .581.

Bradley-Geist et al., 2010 Study 3 xxx	We predicted that incentivized participants would not be credentialed because they would realize that their choice could be interpreted as reflecting their own self-interests as opposed to altruism and authenticity.	2 (valence: positive experience, negative experience) x 2 (choice: forced choice, free choice) x 2 (incentive: avoiding prejudice incentive, avoiding personal failure incentive)	Difference of means	A MANOVA, which included all three independent variables, was run on the dependent variables of job suitability and tolerance for prejudice. Given the theoretically relevant missing data in the Free Choice × Negative Valence conditions, only two interactions were tested: (a) the interaction between the incentive variable (avoiding prejudice, avoiding personal failure) and the valence (positive, negative) variable and (b) the interaction between the incentive (personal failure, prejudice) variable and the choice (forced, free) variable. Given the larger sample size in this study compared to Studies 1 and 2, we were able to enter participant gender and age as covariates to ensure that the conclusions were not affected by these demographic characteristics. The only significant multivariate effect was a two-way interaction between the incentive (personal failure, prejudice) variable and the valence (positive, negative) variable, Wilks's Lambda = .96, $F(2, 130) = 2.9$, $p = .05$. Separate univariate ANOVAs were run on each of the dependent variables. Job suitability was not significantly related to any of the independent variables or their interactions. However, the manipulations did affect tolerance for prejudice; a two-way interaction between the incentive variable and the valence variable emerged, $F(1, 117) = 3.30$, $p = .05$. Planned comparisons suggest that individuals in the avoiding personal failure incentive condition tended to express less tolerance for prejudice when they wrote about a negative experience			The authors do not test their predictions. They make predictions about the effects of incentives but do not have a no-incentives control condition. Surprisingly, the authors report different *M*s and *SD*s for the same condition within the same paragraph. Specifically, 'individuals in the avoiding individual failure incentive condition tended to express less tolerance for prejudice when they wrote about a negative experience ($M = 2.17$, $SD = 0.28$)...' and 'individuals in the avoiding personal failure condition tended to express more tolerance of prejudice ... when they wrote about a negative experience ($M = 2.67$, $SD = 0.63$).

				(*M* = 2.17, *SD* = 0.28) than did individuals in the avoiding prejudice condition who wrote about a negative experience (*M* = 2.75, *SD* = 1.17), *F*(1, 42) = 1.97, *p* < .10. Individuals in the avoiding personal failure condition tended to express more tolerance of prejudice when they wrote about a positive experience (*M* = 2.17, *SD* = 1.20) than when they wrote about a negative experience (*M* = 2.67, *SD* = 0.63), but none of the other condition comparisons was statistically significant.			
Bradley-Geist et al., 2010 Study 4	We anticipated that choosing to write about a positive interaction with a Black friend would result in the highest level of credentialing more so than choosing to write about a positive interaction with a black acquaintance. The choice of writing something positive and the ability to point to a close relationship with a Black person should give participants confidence that they have "proven" that they harbor no racial biases.	2 (choice: forced choice, free choice) x 2 (valence: positive, negative) x 2 (closeness: close friend, acquaintance)	Attenuated two-way interaction	Job suitability: Choice and closeness. There were no significant main effects for choice or closeness, nor was there a significant interaction between choice and closeness on the job suitability measure. Tolerance for prejudice: Choice and closeness. There was no main effect of choice on tolerance for prejudice, nor was there a significant interaction between choice and closeness.			The authors do not find what they predict. They predict a closeness x choice interaction on both dependent measures, but do not find any effects.
Brown et al., 2011 Study 1	We hypothesized that moral credentialing would result in more misconduct when that misconduct is highly rationable. In contrast, moral credentialing should not enhance misconduct when that misconduct is unambiguously immoral, in	2 (moral credentialing: self vs. acquaintance) × 2 (rationalizability:	Attenuated two-way interaction	As predicted, however, the analysis of variance also revealed a Credentialing × Rationalizability interaction, *F*(1, 187) = 3.80, *p* = .05, *MSE* = 9.44.			Participants in the moral credentialing self condition read four moral dilemmas and rated how likely they would be to behave in a prosocial fashion if they were to encounter such a situation. Moral licensing

x xxx	which case it would be difficult to rationalize.	high or low) anal-					theory would predict that only participants who indicate that they would be to behave more prosocial display licensing effects (but not participants who indicate that they would not behave in a prosocial fashion). The authors do not give information on how participants in this condition responded to the scenarios. Theoretically, the licensing effect would thus only occur for people who indicate that they would behave prosocial, but the authors are not clear whether only these people are included in the moral credentialing self condition. When inserting the *F*-value and the degrees of freedom on http://graphpad.com/quickcalcs/pValue2/, we found that the actual *p*-value is not .05, but .0546.
Cain et al., 2005	We test the following three predictions: .... 2) Advisors with conflicts of interest will give more biased advice under conditions	Two-cell repeated measures (disclosur	Difference of means	More interesting, and as predicted, all three measures also reveal that disclosure led to a greater distortion of advice. The amount that advisors exaggerated, calculated by			Exact *p*-values are not reported. When calculating* the *p*-values, we found that the comparisons between

Study 1 xx	with disclosure than without disclosure. ….	e versus not disclosure) but licensing is only tested in 2 condition s. We compare those based on the reported means and SD's.		subtracting advisors' own personal estimates from their public suggestions, was significantly greater in the high/disclosed condition than in either of the other two conditions (p < .05) and significantly greater by the other two measures as well: advisor suggestion minus actual jar values and advisor suggestion minus the average of personal estimates in the accurate condition (p < .05 for both).			the high/undisclosed condition (M = 20.16, SD = 4.81) and the high/disclosed condition (M = 24.16, SD = 8.40) were not significant, t(41) = 1.9275, p = .061 (which is not <.05).
Cain et al., 2010 Study 1	We designed our first study to address the question of whether and shy advisors expect disclosure to shift the peak of the advisee response curve to the left or right.	Two-cell (disclosur e versus no disclosure)	Difference of means	Average advice (number of jellybeans) did not differ significantly between subjects if we compare the first response of those who got the disclosure condition first (M = 4017, SD = 1928) versus the first response of those who got the nondisclosure condition first (M = 4333, SD = 2092). The mean difference in advise (disclosure minus nondisclosure, within subject) was 13 jellybeans (SD = 1479) higher for nondisclosure when the nondisclosure condition came first, and was 1022 (4661) higher for nondisclosure when the disclosure condition came first.			Since the authors do not make clear predictions, we could not include this comparison in our p-curve analysis. Moral licensing implies, in our view, that advisors should give higher estimates of the amount of jelly beans in the jar when their COI is disclosed vs when it is undisclosed. This is not the case.
Cain et al., 2010	(2) As in figure 1, advisors with conflicts of interest will give more biased advice under conditions	Two-cell (disclosur e versus no	Difference of means	Table 1. Advisor's suggestion. High-undisclosed M = 236,138, SD = 36,071, High - disclosed M = 255,394, SD = 55,877.			The authors report a test of the effects of disclosure in the column ' high vs. high-d', and report 'p <. 05'. Our

Study 3 xx	with disclosure than without disclosure…..	disclosure)					recalculation* showed that $t(75) = 1.7686$, $p = .081$
Cascio & Plant, 2015 Study 1	We predicted that agreeing to take part in the figure fundraiser would license our participants to make a racially biased decision in the present (say a job was better for a White than Black candidate).	Two-cell (prospecti ve moral behavior versus control)	Difference of means	Results revealed that participants in the future moral behavior condition were more likely to indicate that the job was better suited for a White candidate ($M = 4.62$, $SE = 0.13$) than control participants ($M = 4.16$, $SE = 0.12$), $t(78.829) = 2.49$, $p = .015$, $d = 0.55$. Thus, intending to perform a moral behavior in a few weeks led to the expression of more racial bias in the moment.	$t(78.829) = 2.49$		
Cascio & Plant, 2015 Study 2	Study 2 was designed to replicate and extend study 1. Participants were once again told about the skip-a-meal fundraiser but instead of having them commit and provide their email address, the experimenter indicated that they were just taking a quick poll of students' interest in the fundraiser. This allowed us to assess whether participants would experience the prospective moral licensing effect when they had only indicated whether they anticipated taking part but had not committed to the fundraiser.	Two-cell (prospecti ve moral behavior versus control)	Difference of means	Results from the *t*-test with equal variances not assumed revealed that, consistent with Study 1, participants in the prospective moral behavior condition expressed greater preference for a White candidate ($M = 4.54$, $SE = 0.20$) than control participants ($M = 4.00$, $SE = 0.12$), $t(38.1) = 2.33$, $p = .025$, $d = 0.68$.	$t(38.1) = 2.33$		

Cascio & Plant, 2015 Study 3 x	Study 3 was designed to test the boundary conditions of prospective moral licensing. First, participants were asked about a different future moral behavior, donating blood, and a different measure of racial prejudice was used, responses on the Attitudes Towards Blacks scale (Brigham, 1993). Second, an additional control condition was added to the design in which the participants were asked to perform a future non-moral behavior. This way, we ruled out the possibility that agreeing to any request, rather than a specifically moral re- quest, results in more racially biased responding. Finally, we were also interested in determining whether prospective moral licensing occurs for overt as well as ambiguous prejudice-relevant behaviors. The behavior assessed in both Study 1 and Study 2 was relatively ambiguous, and participants may have interpreted their responses as not being prejudiced (e.g., perceived the job as likely unpleasant for a Black officer). Examining whether the prospective moral licensing effect extends to responses to overly racially biased items on a self-report measure will provide	Three cells, two control conditions	Difference of means moral vs control in main test, robustness: moral vs nonmoral	A one-way 3 (Future Moral vs. Future Nonmoral vs Control) ANOVA was conducted on the overt prejudice scale. The analysis revealed that condition significantly affected the amount of prejudice expressed, $F(2, 78) = 12.136$, $p < .001$, $\eta^2 = .07$ (see Fig. 1). Planned contrasts were then used to compare the future moral licensing condition to each of the other conditions. Participants in the future moral condition ($M = 2.87$ $SE = 0.23$) expressed significantly more prejudice than participants in the non-moral condition ($M = 1.43$, $SE = 0.21$) and than control participants ($M = 1.68$, $SE = 0.22$), ps < .001, ds = 1.14 and 0.89, respectively. Participants in the non-moral condition and control condition did not differ from each other, $p = .42$. It is worth noting that if the full 20-itemversion of the ATB is used, the pattern of results is similar, although weaker, indicating that the effect tends to be stronger for the more blatant items.			Not included because licensing test is not reported. Recalculations using http://www.graphpad.com/quickcalcs/ttest1/?Format=SEM showed that $t(50) = 3.7398$, $p < .001$ and $t(52) = 4.6294$, $p < .001$.

	insight into whether the licensing is due to moral credentials being established or moral credits accumulating. If the moral licensing effect extends to the overt items on the measure of racial prejudice, it suggests that moral credits rather than moral credentials are at play in prospective moral licensing						
Cascio & Plant, 2015 Study 4	The present study was designed to provide a replication of moral request vs. non-moral request conditions and provide further evidence that prospective moral licensing is due to moral credits rather than moral credentials. Participants were asked about either participating in a future fundraiser or using a new email system, and then they completed an explicit measure of racial bias, a stereotype endorsement measure. The endorsement of negative stereotypes of Black people rep- resents overt unambiguous racial bias. If the moral licensing effect ex- tends to overt stereotype endorsement, it provides further evidence that prospective moral licensing is due to moral credits. Method	Two-cell (prospecti ve moral behavior versus control)	Difference of means	Results from the *t*-test with equal variances not assumed revealed that participants in the prospective moral behavior condition expressed stronger stereotype endorsement (M = 3.84, SE = 0.11) than non-moral request participants (M = 3.39, SE = 0.14), $t(69.34) = 2.57$, $p = .012$, $d = 0.59$.	*t*(69.34) = 2.57		

Choi et al., 2014 Study 2	Hypothesis 1. People who first rate a high-quality Black-modeled ad (legitimacy condition) will discriminate against a subsequent low-quality Black-modeled ad more than those who first rate a high-quality White-modeled ad (suppression condition).	Two-cell (first high quality black versus white)	Difference of means	Participants who first evaluated a high-quality ad with a Black and then a low-quality ad with a Black expressed a significant bias against the target ad compared with those who first evaluated a high-quality ad with a White and then a low-quality ad with a Black (*M*black/black = 4.15 vs *M*white/black = 4.77, $t = 2.35$, $p < .05$). These results support Hypothesis 1.	$t(115) = 2.35$		*SD*s not reported, *n* not reported, so we cannot be sure of *df*.
Choi et al., 2014 Study 2	Also, related to hypothesis 1, we predicted that discrimination would appear against the low-quality ad containing Black model compared with the low-quality ad with White model when subjects were asked to evaluate a high-quality ad featuring the model first.	Two-cell (first high quality black versus white)	Difference of means	When subjects gained legitimacy credits through the process of evaluating a high-quality ad featuring Black model positively, they subsequently showed a fairly strong and significant bias against the low-quality ad containing Black model as compared with the low-quality ad with White model (*M*black/black = 4.15 vs *M*black/white = 5.16, $t = 3.49$, $p < .05$). This pattern of data showed strong support for the legitimacy credit interpretation.	$t(115) = 3.49$		*SD*s not reported, *n* not reported, so we cannot be sure of *df*.
Choi et al., 2014 Study 3	Hypothesis 2. People who acquire legitimacy credits will be more likely to spend them under the condition where such spending is sensibly justified than under the condition where such spending is not. That is, those who acquire legitimacy credits will be more likely to devaluate an ad with Black only when the ad quality is low and hence, such discrimination is normatively appropriate.	Two-cell (first Mno versus Mhigh)	Difference of means	While there was a significant difference between control (low) and second ad of low-quality condition (*M*no/low = 3.60 vs *M*high/low = 3.03, $t = 2.11$, $p < .05$). This suggests that people are likely to display prejudice expression only when it is normatively appropriate to do so even after they acquire legitimacy credits. The results support Hypothesis 2.	$t(58) = 2.11$		*SD*s not reported, *n* not reported, so we cannot be sure of *df*.

Conway & Peetz, 2012 Study 1 x	We expected more prosocial intentions if participants recalled a distant moral behavior than distant immoral behavior (a consistency effect) and more prosocial intentions if participants recalled a recent immoral behavior than a recent moral behavior (a compensatory effect).	2 (event valence: moral vs immoral) x 2 (event distance: recent versus distant)	Reversing interaction	Willingness to volunteer. As predicted, there was a significant interaction, $F(1, 96) = 5.37$, $p = 0.23$, $\eta^2 = .05$.			The licensing comparison is not made because it is not explicitly predicted. However, the authors write about moral licensing in the introduction: In sum, there is ample evidence showing that, sometimes, perceiving oneself as moral leads to a reduction in moral behavior, whereas perceiving oneself as less moral motivates increases in moral behavior. This pattern is interpreted in terms of moral self-regulation, balancing moral self-consistent behavior against the costs inherent in prosocial behavior. It is furthermore predicted that when people recall their behavior in a manner that induces a concrete mind-set (e.g., recalling recent behavior), they should demonstrate compensatory effects: Feeling moral should reduce prosocial intentions, whereas feeling immoral should increase prosocial intentions. Therefore, we tested* the cells in which licensing is predicted (recent moral versus recent immoral

							conditions), and found for the willingness to volunteer scales that participants in the recent moral condition (M = 5.42 ,SD = 1.37) do not differ significantly from participants in the recent immoral condition (M = 5.97, SD = 1.30), t(49) = 1.46, p = .150
Conway & Peetz, 2012 Study 1 x	We expected more prosocial intentions if participants recalled a distant moral behavior than distant immoral behavior (a consistency effect) and more prosocial intentions if participants recalled a recent immoral behavior than a recent moral behavior (a compensatory effect).	2 (event valence: moral vs immoral) x 2 (event distance: recent versus distant)	Reversing interaction	Willingness to help. As predicted, there was, again, a significant interaction, F(1,96) = 5.97, p = 0.16.			The licensing comparison is not made because it is not explicitly predicted. However, the authors write about moral licensing in the introduction, see previous comment. Therefore, we tested* the cells in which licensing is predicted (recent moral versus recent immoral conditions), and found for the willingness to help people in the hypothetical vignettes that participants in the recent moral condition (M = 5.81, SD = 0.70) do not differ significantly from participants in the recent immoral condition (M = 6.00, SD = 0.73), t(49) = 0.94, p = .353.

Conway & Peetz, 2012 Study 2 x	In the next study, we aim to establish that the interaction obtained in Study 1 was due to changes in moral self-perceptions, rather than simply greater activation of moral values in the distant condition. We expected that compensatory and consistency patterns of subsequent prosocial motivation would occur only for self-relevant past behavior, not for moral behavior by others.	2 (actor: self versus other) x 2 (event valence: moral vs immoral) x 2 (event distance: recent vs distant)	Attenuation of reversing interaction (reversing interaction for self vs none for other.)	As predicted, results also revealed a significant three-way interaction, $F(1,143) = 5.48$, $p = .021$, $\eta^2 = .04$.			The licensing comparison is not made because it is not explicitly predicted. However, the authors write about moral licensing in the introduction, see previous comment. Therefore, we tested* the cells in which licensing is predicted (recent moral versus recent immoral conditions), and found that participants in the moral condition ($M = 4.66$, $SD = 1.53$) did not differ from participants in the immoral condition ($M = 5.51$, $SD = 1.34$), $t(37) = 1.85$, $p = .073$.
Conway & Peetz, 2012 Study 3	We expected compensatory behavior patterns if participants were instructed to consider a concrete behavior and expected consistency behavior patterns if participants were instructed to consider the actions' implications for identity.	2 (moral content: moral versus immoral) x 2 (focus: identity vs action)	Reversing interaction	As predicted the main effect of valence was qualified by a significant interaction, $F = 4.06$, $p = .046$, $\eta^2 = .04$.			Not included in *p*-curve because *df*'s are not reported (while they are in the other experiments).
Cornelissen et al., 2013a	In this study, we tested our hypothesis that a consequential mind-set leads to moral balancing, whereas a deontological mind-set results in moral consistency. We expected	Two cell (only outcome based conditions: ethical	Difference of means	Participants with an outcome based mind set gave less coins after recalling an ethical act ($M = 2.37$, $SD = 2.31$) than after recalling an unethical act ($M = 3.71$, $SD = 2.13$), $t(46) = -2.07$, $p = .04$	$t(46) = -2.07$		

Study 1	participants identified as having a rule-based mind-set to use the ethicality of their recalled behavior as a guide, such that those who recalled an ethical behavior would behave more fairly than those who recalled an unethical act. For participants with an outcome-based mind-set, we expected the opposite effect.	vs unethical)					
Cornelissen et al., 2013a Study 2	The objective of the second study was to provide additional evidence for the causal role of ethical mind-sets as a moderator of moral dynamics by manipulating instead of measuring them.	Two cell (only outcome based conditions: ethical vs unethical)	Difference of means	In the outcome based mind set condition, participants who recalled an unethical act gave more coins in the dictator game (M = 3.20, SD = 1.91) than did those who recalled an ethical act (M = 1.86, SD =1.96), $t(39) = -2.22$, $p = .03$.	$t(39) = -2.22$		
Cornelissen et al., 2013a Study 3 xx	After replication our hypothesized moderation effect, we wanted to evaluate the generalizability of our findings with a third study, so we changed the context to cheating behavior. Additionally, we evaluated the role of moral self-image as the underlying mechanism for moral-balancing effects.	Two cell (only outcome based conditions: ethical vs unethical)	Difference of means	In the outcome based mind set condition, participants who recalled an ethical act (M = 1.74, SD = 2.03) cheated more than did participants who recalled an unethical act (M = 0.74, SD =1.86), $t(52) = 1.97$, $p = .05$.	$t(52) = 1.97$		When inserting the t-value and the degrees of freedom on http://graphpad.com/quickcalcs/pValue2/, we found that the actual p-value is > .05 (.0542)
Cornelissen et al., 2013b	We expect that moral licensing effects of symbolic actions will be especially pronounced for individuals that are highly	Two cell (symbolic	Difference of means	We found a main effect of the symbolic action $t(67) = -2.40$, $p < .03$.	$t(67) = -2.40$		"Published" as symposium paper, so included as published in this p-curve.

Study 1	concerned with impression management. In study 1, we tested whether having the opportunity to engage in a symbolic action leads to a licensing effect.	action vs control)					
Cornelissen et al., 2013b Study 2 x	Study 2 was designed to replicate the moral licensing effect of symbolic actions using a different dependent variable.	Two cell (symbolic action vs control)	Difference of means	The interaction effect of self-monitoring and having the opportunity to perform a symbolic action was not significant ($t(114) = -1.41$, $p = .16$. For participants high in self-monitoring (at 1 *SD* above the mean) there was a marginally significant effect of symbolic action (90% CI [-1.00; -0.05]).			The authors do not provide the test for the licensing effect and we could not compute it from the reported data. Figure 2 and the CI's on page 10 suggest that the licensing effect is not present in the data.
Effron et al., 2009 Study 1	We propose that endorsing Obama can license people to favor whites at the expense of blacks. Study 1 tests whether letting participants endorse Obama increases their willingness to favor a White job applicant.	Two cell (credentials versus control condition)	Difference of means	As predicted, participants who completed the police-hiring task after endorsing Obama said the job was significantly better for a white person ($M = .55$, $SD = .15$) than did control participants ($M = .17$, $SD = .12$), $t(82) = 2.04$, $p = .04$, $d = .45$.	$t(82) = 2.04$		
Effron et al., 2009 Study 2	Simply expressing one's preference for a democrat may have been sufficient to produce the effect (political expression account). Second, perhaps seeing Obama activated stereotypes about blacks that biased participants' decision towards a white applicant (priming	Three cells (credentials vs political expression control versus	Difference of means	Planned orthogonal contrasts confirmed our predictions that while responses in the political expression control ($M = .05$, $SD = .84$) and the priming control ($M = .15$, $SD = 1.09$) did not differ significantly from each other, $F(1,59) = .13$, ns, participants in the credentials condition ($M = .50$, $SD = .83$) favored Whites for the job significantly more			Recalculating* *t*-tests for each comparison between control 1, 2, and treatment with assumed $n = 25$ per condition even though total $N = 74$ showed that both comparisons are not significant: $t(48) = 1.91$, $p = .063$ for the political

xx	account). We conducted study 2 to rule out these alternatives.	priming control)		than participants in the two control condition did $F(1,59) = 5.65$, $p = .02$			expression control condition, and $t(48) = 1.28$, $p = .208$ for the priming control condition.
Effron et al., 2009 Study 3	If endorsing Obama licenses favoring whites, then it should have an especially strong effect on individuals whose preexisting attitudes dispose them towards white favoritism, and who should thus be most inhibited in the absence of credentials. We tested this prediction in study 3 by including a standard measure of prejudice: the modern racism scale. We predicted that the participants with higher MRS scores would be more likely to take advantage of credentials to favor whites. We predicted that endorsing Obama would increase allocations to the white organization, especially among participants high in modern racism.	2 (credentials vs political expression condition) x 2 (high mrs vs low mrs)	Attenuated two-way interaction	At 1 *SD* above the mrs mean, we replicated the moral credentials effect: the white organization received $16.478 more (and the black organization received $ 16.478 less) in the credentials condition than in the control condition, $t(48) = 2.48$, $p = .02$. As predicted, this effect was weaker at the MRS mean: credentials increased the white organization's allocation by $2,555, $t(48) = .57$.			Authors do not report the full licensing test, so we stayed as close to the meta analysis as possible (and looked at the effect at the MRS mean), $t(48) = 0.57$, n.s.
Effron et al., 2012 Study 3a	We hypothesized that participants who could point to a racist alternative to their prior behavior would be more likely to express preferences that favored whites at the expense of blacks.	Two cell (innocent black suspect versus control.)	Difference of means	Consistent with this hypothesis, participants expressed a marginally stronger preference for hiring whites in the innocent black suspect condition ($M = .29$, $SD = .72$) than in the control condition ($M = .11$, $SD = .56$), $t(155) = 1.77$, $p < .08$, $d = .028$.			n.s.

Effron et al., 2012 Study 3b	We hypothesized that participants who could point to a racist alternative to their prior behavior would be more likely to express preferences that favored whites at the expense of blacks.	Two cell (innocent black suspect versus control.)	Difference of means	Although participants tended to allocate a slightly greater percentage of the money to the white organization in the innocent black suspect condition (M = 54.45%, SD = 11.31) than in the control condition (M = 52.226%, SD = 10.79), this difference was not significant, $t(155) = 1.28$, $p = .20$, $d = .21$).			n.s.
Jordan et al., 2011 Study 2	We predicted that, in comparison to control participants, recalling immoral behaviors would increase prosocial intentions and recalling moral behaviors would decrease prosocial intentions.	Three cells (moral recall vs immoral recall vs control)	Difference of means	As predicted, relative to the control, recalling moral or immoral behavior led to compensatory prosocial intentions, but remembering a positive or negative nonmoral behavior did not. A one-way anova on participants' prosocial intentions revealed a significant effect of condition, $F(4,143) = 3.46$, $p = .01$, $\eta^2 = .09$. Relative to controls, participants in the moral condition had marginally weaker prosocial intentions $F(1, 143) = 3.36$, $p = .069$, $\eta^2 = .05$			Study actually has 5 cells. The authors only report overall ANOVA on 5 cells and two separate *f*-tests for control <> moral and control <>immoral. The licensing test = control > moral. Which is n.s. $F(1.143) = 3.36$
Jordan et al., 2011 Study 3a	We predicted that recalling one's own immoral behavior would lead to less cheating (and more persistence) than recalling one's own moral behavior.	2 (self other) x 2 (moral immoral)	Difference of means	A comparison of the self stories indicated that, after recalling moral behavior, participants allowed the answer to appear more than they did after recalling immoral behavior $t(44.26) = 4.29$, $p < .001$, $\eta p^2 = .21$	$t(44.26) = 4.29$	$t(44.89) = 42.88$ $t(70.43) = 4.34$	
Khan & Dhar, 2006 Study 1	We predict that participants will be more likely to choose the luxury option when their prior decision provides them with an opportunity to appear altruistic by committing to a virtuous act	Two cell (license versus control condition)	Difference of proportion	Consistent with our prediction, significantly more people in the license condition chose the designer jeans (57.4%) than in the control condition (27.7%, x2 = 9.7, $p < .01$)	chi2(1) = 9.7		

Khan & Dhar, 2006 Study 2a	We predict that participants will be more likely to choose the luxury option when their prior decision provides them with an opportunity to appear altruistic by committing to a virtuous act	Two cell (license versus control condition)	Difference of proportion	As we predicted, 56.6% of participants chose to buy the expensive sunglasses in the license condition; only 27.7% of participants in the control condition chose the expensive sunglasses (x2 = 7.95, $p < .05$)	chi2(1) = 7.95		
Khan & Dhar, 2006 Study 2b	We predict that participants will be more likely to choose the luxury option when their prior decision provides them with an opportunity to appear altruistic by committing to a virtuous act	Two cell (license versus control condition)	Difference of proportion	The choice results were similar to the previous studies: 45% of the participants in the control condition chose to buy the concert ticket, whereas 72.5% of the participants in the license condition opted to buy the ticket (x2 = 6.24, $p < .05$)	chi2(1) =6.24		
Khan & Dhar, 2006 Study 3	we predicted that, on average, participants would donate a smaller amount of money if they felt licensed by their prior decision than would participants in the control condition.	Two cell (license versus control condition)	Difference of means	However, consistent with our prediction, among those who donated, donations were significantly lower in the licensing condition (M = \$1.20) than in the control condition (M = \$1.70; t (70) = 2.73, $p < .01$	$t(70)$ = 2.73		
Khan & Dhar, 2006 Study 4	We predict that participants will be more likely to choose the luxury option when their prior decision provides them with an opportunity to appear altruistic by committing to a virtuous act	Two cell (license versus control condition)	Difference of proportion	We replicated the licensing effect. That is, compared with the control condition (40%), significantly more participants opted for the designer jeans in the licensing condition when there was no external attribution (62%; x2 = 4.05, $p < .05$)	chi2(1) = 4.05		
Khan & Dhar, 2006	We predict that participants will be more likely to choose the luxury option when their prior decision provides them with an	Two cell (license versus	Difference of means	The relative preference for the hedonic designer jeans was significantly higher in the license condition (M = 5.9) than in the control condition (M = 5.4, $t(64)$ = 2.27, $p < .05$).	$t(64)$ = 2.27		

Study 5	opportunity to appear altruistic by committing to a virtuous act	control condition)					
Mann & Kawakami, 2012 Study 1	Specifically, in Study 1 we investigated the impact that perceived progress on the goal to be more egalitarian toward Blacks had on outgroup discrimination and prejudice. We proposed that after receiving feedback that they were progressing on this goal, participants would disengage from this goal by distancing them- selves from Blacks and by holding more negative implicit racial attitudes than would participants in a no-progress condition.	Two cell (progress vs no progress)	Difference of means	As predicted, participants who were given feedback that they were progressing on the goal to be positive toward Blacks (M = 7.90, SD = 1.18) sat significantly farther away from the Black confederate than did participants in the no-progress condition (M = 6.86, SD = 1.93), $t(41) = 2.12$, $p = .04$, $d = 0.65$.	$t(41)$ = 2.12		
Mann & Kawakami, 2012 Study 2	We therefore predicted that only feedback on goal progress toward Blacks would increase seating proximity to a White confederate and increase negative implicit racial attitudes	2 (goal feedback: progress vs no progress) x 2 (type of goal: black vs white)	Difference of means	This effect, however, was qualified by the predicted Goal Feedback x Type of Goal interaction, $F(1,287) = 5.91$, p =.02, η^2 =.06. As expected, when examining goals related to Blacks, simple effects analyses demonstrated that participants who received feedback that they were becoming more positive toward Blacks subsequently sat closer to the White con- federate (M = 5.41, SD = 1.47) than did participants in the no-progress condition (M = 6.30, SD = 1.03), $t(40) = 2.25$, $p = .03$, d = 0.70.	$t(40)$=2.25		Odd (implicit) reasoning: Sitting close to a white person supposedly signals something about what you think about blacks.

Mann & Kawakami, 2012 Study 3	We predicted that feedback on goal progress toward Blacks would decrease seating proximity to a Black confederate and increase seating proximity to a White confederate relative to the no- progress and no-feedback conditions.	3 (goal feedback: progress vs no progress vs no feedback) x 2 (target race: black confederate vs white confederate).	Difference of means	To examine the effect that feedback related to progress on the goal to be positive toward Blacks had on nonverbal behavior toward Whites and Blacks, we performed a Goal Feedback (progress vs. no progress vs. no feedback) x Target Race (Black confederate vs. White confederate) ANOVA on seating distance. Only the Goal Feedback x Target Race interaction was significant, $F(2, 80) = 9.53$, $p = .001$, $\eta^2 = .19$. Simple effects analyses related to the Black confederate replicated the findings in Study 1, $F(2, 40) = 4.82$, $p = .01$, $\eta^2 = .21$. In particular, participants who received feedback that they were becoming more positive toward Blacks sat farther from the Black confederate ($M = 43.78$, $SD = 1.81$) than did participants in the no-progress condition ($M = 40.38$, $SD = 5.13$), $t(27) = 2.41$, $p = .02$, $d = 0.88$, and participants in the no-feedback condition ($M = 37.66$, $SD = 7.46$), $t(25) = 3.08$, $p = .005$, $d = 1.13$. Participants in the no-progress feedback condition, alternatively, did not differ in seating distance to the Black confederate ($M = 40.38$, $SD = 5.13$) from participants in the no-feedback condition ($M = 37.66$, $SD = 7.46$), $t(24) = 1.10$, $p = .28$, $d = 0.42$.	$t(27)=2.41$	$t(25)=3.08$	Enormous effect size.
Mazar & Zhong, 2010	On the basis of research on behavioral priming, we predicted that mere exposure to green products would increase subsequent altruistic conduct; however, on the basis of recent	2 (store: green vs conventional) x (behavior: exposure	Attenuated two-way interaction	However, there was a significant interaction, $F(1, 152) = 4.45$, $p = .037$, $p_{rep} = .897$. Participants who were merely exposed to the green store shared more money ($M = \$2.12$, $SD = \$1.40$) than those who were merely exposed to the conventional store ($M = \$1.59$,			Test of interest is ' purchased in green store" but is n.s. $F(1,152) = 1.69$

Study 2	theories on moral regulation, we predicted that purchasing green products would reduce subsequent altruism because it establishes moral credentials. Experiment	versus purchasing)		SD = $1.29), *F*(1, 152) = 2.85, *p* = .094, *p*rep = .824. This pattern reversed in the purchasing conditions: Participants who had purchased in the green store shared less money (*M* = $1.76, *SD* = $1.40) than those who had purchased in the conventional store (*M* = $2.18, *SD* = $1.54), *F*(1, 152) = 1.69, *p* = .195, *p*rep = .728.			
Mazar & Zhong, 2010 Study 3	We predicted that purchasing green products would establish moral credentials, ironically licensing selfish and morally questionable behavior.	Two cell (license versus control condition)	Difference of means	We found a significant difference between conditions in performance on the dots task, *t*(79) = 2.26, *p* = .027, *p*rep = .913. Participants who had purchased in the conventional store identified 42.5% (SD = 2.9%) of trials as having more dots on the right side; this percentage was not significantly different from the actual percentage (i.e., 40%), *t*(37) = 1.66, *p* = .106, *p*rep = .811. Participants who had purchased in the green store, however, identified 51.4% (*SD* = 2.67%) of trials as having more dots on the right side -which suggests that they were lying to earn more money. Participants in the green-store condition earned on average $0.36 more than those in the conventional-store condition. As noted, independently of deciding to lie, participants could steal by taking more money from the envelope than shown on the summary screen. Results for this measure were consistent with those for task performance: Participants in the green-store condition stole $0.48 more from the envelope than those in the conventional-store condition (*M* = $0.56, *SD* = $0.13, vs. *M* = $0.08, *SD* = $0.14), *t*(79) = 2.55, *p* = .013, *p*rep	*t*(79) = 2.26	*t*(79) = 2.55	

				= .942. Altogether, participants in the green-store condition left the experiment with on average $0.83 (*SD* = $0.23) more in their pockets than did participants in the conventional-store condition, *t*(70) = 3.55, p < .001, prep > .986.			
Merritt et al., 2012 Study 2 follow up	We have argued that participants strategically expressed racial sensitivity in order to reduce their concern that rejecting a Black candidate would seem prejudiced.	Two cell (license versus control condition)	Difference of means	Yet, as predicted, this preference was stronger when participants had had an opportunity to establish credentials, (*M* = 7.42, *SD* = 1.11) than when they had not (*M* = 6.75, *SD* = 1.58), *t*(67) = 2.03, *p* = .05.	*t*(67)=2.03		
Monin & Miller, 2001 Study 1	Each of the present three studies tests the hypothesis that people who have previously expressed antiprejudiced attitudes are more likely to express their true attitudes in contexts in which there exists the potential for accusations of prejudice. We hypothesized that participants would be more likely to explicitly express an implicit preference for a man when their previous actions provided them with non- sexist credentials.	2 (gender: man vs woman) x 3 (condition : most vs some vs base-rate)	Difference of means	Effect of credentials. We hypothesized that participants in the "most" condition would be more likely than would those in the "some" condition to deviate from the politically correct midpoint option and express a preference for hiring a man. To test this hypothesis, we analyzed participants' preferences in a 2 X 3 factorial ANOVA, crossing gender with condition (most, some, and base rate). Consistent with predictions, this analysis yielded a main effect for credentials, *F*(2, 194) = 4.4, *p* < .05. As Figure 1 illustrates, the main effect for credentials reflects the fact that participants favored a man more in the "most" condition (*M* = 4.8) than they did in either the "some" or the base-rate condition (*M*s = 4.3 and 4.5, respectively).	*F*(2,194) = 4.4		Unexpectedly, they only find the effect for men. Still, prediction was overall, therefore we included only the main effect of credentials.

Monin & Miller, 2001 Study 2	Each of the present three studies tests the hypothesis that people who have previously expressed antiprejudiced attitudes are more likely to express their true attitudes in contexts in which there exists the potential for accusations of prejudice. We hypothesized that the act of recommending the hiring of a woman [an African American] on one job selection task would increase participants' willingness to express the belief that a man [a White] was more suited for a position described on a second job selection task. 1	Two cell (credentials versus control condition)	Difference of means	We conducted a 2 (credentials) X 2 (type of prejudice) X 2 (gender) ANOVA on the responses to the job suitability question. The only significant effect to emerge was the predicted main effect for credentials, F(l, 124) = 6.3, p < .05, all other Fs n.s. As Figure 2 shows, participants were more likely to favor a White man in the credentials condition (M = 4.8) than in the noncredentials condition (M = 4.4). The fact that credentials did not interact with type of prejudice, F(l, 124) = 0.2, ns, indicates that the manipulation had comparable effects on the racism and sexism measures.	F(1,124) = 6.3		
Monin & Miller, 2001 Study 3	Each of the present three studies tests the hypothesis that people who have previously expressed antiprejudiced attitudes are more likely to express their true attitudes in contexts in which there exists the potential for accusations of prejudice.	Two cell (credentials versus control condition)	Difference of means	Expressed prejudice. To begin with, we checked to see whether the findings of Study 3 replicated those of previous studies. We did this by comparing the credentials and no-credentials conditions under the same-audience condition. The earlier findings were replicated in that participants who were given the opportunity to establish nonracist credentials favored a White more than did those who were not given the opportunity to establish credentials (M = 4.9 vs. 4.2), F(19) = 2.2, p < .05.	t(19) = 2.2		
Sachdeva et al., 2009	Our primary hypothesis was that a decrease in the moral self-concept leads to increased prosocial behavior, but that an	Three cell (positive recall, negative	Difference of means	We compared the average donation across conditions using a one-way analysis of variance (ANOVA). Participants in the neutral condition donated an average of			Authors do not report the critical licensing test. When calculating* the comparison between the positive traits

Study 1 x	analogous increase in the moral self-concept inhibits altruistic or prosocial behavior (relative to a neutral state). We also tested our conjecture that these effects are caused by changes in the self-concept and therefore should occur when participants are taking a first-person perspective, but not when they are taking a third-person perspective.	recall, control)		\$2.71 (*SE* = 0.85). However, the amount donated increased to \$5.30 (*SE* = 1.17) in the negative- traits condition and fell to \$1.07 (*SE* = 0.47) when participants were primed with the positive traits, $F(2, 43) = 5.690$, *p*rep = .97, $\eta^2 = 209$.			(*M* = 1.07, *SD* = 1.82) and the control condition (*M*= 2.71, *SD* = 3.29), we found $t(27) = 1.68$, $p = .105$
Sachdeva et al., 2009 Study 2 x xxx	Our primary hypothesis was that a decrease in the moral self-concept leads to increased prosocial behavior, but that an analogous increase in the moral self-concept inhibits altruistic or prosocial behavior (relative to a neutral state). We also tested our conjecture that these effects are caused by changes in the self-concept and therefore should occur when participants are taking a first-person perspective, but not when they are taking a third-person perspective.	2 (condition: positive vs negative traits) x 2 (perspective: self vs other)	Attenuated two-way interaction	Once again, the dependent variable was the amount participants indicated they would donate. We ran a 2 (condition: positive or negative traits) x 2 (perspective: self or other) between-subjects ANOVA on this measure. A significant interaction between condition and perspective revealed that the effect of the different word types was apparent only in participants who were told to write about themselves, $F(1, 35) = 4.73$, *p*rep = .90, $\eta^2 = .119$. The donations of participants who wrote about someone they knew did not differ depending on whether they wrote about positive or negative traits (*M* = \$4.55, *SE* = 1.42, vs. *M* = \$3.30, *SE* = 1.27), but participants who wrote only about themselves gave significantly less when they had written a story containing the positive, rather than the negative, traits (*M* = \$1.11, *SE* =0.73, vs. *M* = \$5.56, *SE* =1.55			Authors do not report the critical licensing test. When calculating* the comparison between the positive traits (*M* = 1.11, *SD* = 2.31) and the negative traits (*M* = 5.56, *SD* = 4.9), we found $t(16) = 2.46$, $p = .025$ This study has less than 10 people in each condition to test an interaction. This is really insufficient.
Sachdeva et al., 2009	Our primary hypothesis was that a decrease in the moral self-concept leads to increased	Three cell (positive recall,	Difference of means	Those in the negative-traits conditions chose to run the filters 73% of the time, but those who wrote about the positive traits were the	$F(2,43) = 3.59$		

Study 3	prosocial behavior, but that an analogous increase in the moral self-concept inhibits altruistic or prosocial behavior (relative to a neutral state). We also tested our conjecture that these effects are caused by changes in the self-concept and therefore should occur when participants are taking a first-person perspective, but not when they are taking a third-person perspective.	negative recall, control)		least cooperative and said they would run the filters only 56% of the time, $F(2, 43) = 3.59$, $prep = .90$, $\eta^2 = .143$			
Susewind & Hoelzl, 2014 Study 1 xxx	Applying the rationale of perspectives to the moral domain, we assume that different perspectives people take on their past moral behavior or nonmoral behavior lead to different kinds of feedback, which, in turn, yields different courses of future moral action. Thus, taking a perspective of commitment will lead to perceived commitment in moral or nonmoral behavior and promote a pattern of highlighting. Conversely, taking a perspective of progress will lead to perceived progress in moral or nonmoral behavior and promote a pattern of balancing.	2 (consumption: sustainable vs conventional) x 3 (perspective: unfocused vs progress focused vs commitment focused)	Attenuated two-way interaction	Conversely, in the progress focused conditions, participants exhibited intentions in line with a moral balancing effect, with lower prosocial intentions in the sustainable condition ($M = 3.77$) compared to the conventional condition, $M = 4.10$; Contrast -0.34, $t(237) = -1.80$, $p = .04$ one-tailed, $d = 0.38$.	$t(237)= -1.80$		The authors only provide one-tailed test. When re-calculating the two-tailed test via http://graphpad.com/quickcalcs/PValue1.cfm, we found that $p = 0.0731$
Susewind & Hoelzl, 2014	Applying the rationale of perspectives to the moral domain, we assume that different perspectives people take on their	2 (behavior: moral vs nonmoral	Attenuated two-way	Conversely, participants who perceived their behavior in terms of progress showed a moral balancing effect, offering less money in the moral condition ($M = 1.86$) than in the	$t(122)= -1.75$		The authors only provide one-tailed test. When re-calculating the two-tailed test via

Study 2 xxx	past moral behavior or nonmoral behavior lead to different kinds of feedback, which, in turn, yields different courses of future moral action. Thus, taking a perspective of commitment will lead to perceived commitment in moral or nonmoral behavior and promote a pattern of highlighting. Conversely, taking a perspective of progress will lead to perceived progress in moral or nonmoral behavior and promote a pattern of balancing.	) × 2 (perspective: progress vs commitment)	interaction	non-moral condition, M = 2.43; contrast = -0.57, $t(122)$ = -1.75, p =.041 one tailed, d = 0.43.			http://graphpad.com/quickcalcs/PValue1.cfm, we found that p = 0.0826.
Young et al., 2012 Study 1							Since the authors do not make clear predictions, we could not include this comparison in our p-curve analysis. They describe that the account of reinforcement will predict that recalling good deeds will lead to more good behavior, whereas the account of licensing will predict that recalling good deeds will lead to subsequent bad behavior.
Zhong et al., 2009							The authors do not provide the statistics of the different comparisons. Therefore, we were not able to include their p-value in the table.

Study 1 x							

Note. x = the crucial licensing test was not reported; xx = recalculation shows that test that were reported as significant were not significant at the p < .05 level; xxx = these comparisons included statistical oddities (see comments section). * = *p*-values were (re)calculated from *M*s and *SD*s via http://www.graphpad.com/ quickcalcs/ttest1.cfm?Format=SD

CHAPTER 4

Three Attempts to Replicate the Moral Licensing Effect

The present chapter includes three attempts to replicate the moral licensing effect by Sachdeva, Iliev, and Medin (2009). The original authors found that writing about positive traits led to lower donations to charity and decreased cooperative behavior. The first two replication attempts (student samples, 95% power based on the initial findings, $N_{Study4.1}$ = 105, $N_{Study42}$ = 150), did not confirm the original results. The third replication attempt (MTurk sample, 95% power based on a meta-analysis on self-licensing, N = 940) also did not confirm the moral licensing effect. We conclude that 1) there is as of yet no strong support for the moral self-regulation framework proposed in Sachdeva et al., 2) the manipulation used is unlikely to induce moral licensing, and 3) studies on moral licensing should use a neutral control condition.

This chapter is based on Blanken, I., Van de Ven, N., Zeelenberg, M., & Meijers, M. H. C. (2014). Three attempts to replicate the moral licensing effect. *Social Psychology, 45*, 232-238. doi: 10.1027/1864-9335/a000189

People like to present themselves as good people, both to themselves and to others, to maintain a positive self-image and to feel like a moral person (Aronson, Cohen, & Nail, 1999; Schlenker, 1980; Steele, 1988). Furthermore, central theories of human behavior highlight humans' desire for cognitive consistency in their thoughts, feelings, and behavior (Heider, 1946; Festinger, 1957). Intriguing research on *moral licensing* qualifies this desire for consistency by suggesting that individuals who behave in a morally laudable way, later feel more justified to perform a morally questionable action (Merritt, Effron, & Monin, 2010; Miller & Effron, 2010). Moral licensing is found to lead to a broad spectrum of undesirable behaviors. For example, after (reminders of) prior moral or socially desirable behavior people displayed more prejudiced attitudes (Effron, Cameron, & Monin, 2009; Monin & Miller, 2001), cheated more (Jordan, Mullen, & Murninghan, 2011; Mazar & Zhong, 2010), and displayed a preference for hedonic over utilitarian products (Khan & Dhar, 2006).

An important contribution to the literature on moral licensing examines how writing about one's own positive or negative traits can influence donations to charity and cooperative behavior in a commons dilemma (Sachdeva, Iliev, & Medin, 2009). In just four years since publication, this paper has been cited 129 times (Google Scholar, 11/27/2013). Based on their findings, the authors argued that this moral licensing effect can best be interpreted as part of a larger moral self-regulation framework where internal balancing of moral self-worth and the costs associated with prosocial behavior determine whether one will display (im)moral behavior. When the moral image of oneself is established, an immoral action is allowed without the fear of losing that moral image (moral licensing). However, when one appears immoral to others, positive actions are needed to restore the moral image (moral cleansing). The studies of Sachdeva et al. (2009) comparing licensing with neutral control conditions show medium-sized effect sizes (d = 0.62 ([CL_{95}] -0.11 to 1.35) for Study 1 and d = 0.59 ([CL_{95}] -0.12 to 1.30) for Study 3)[12]. However, note that because of the small sample

[12] Note that the overall differences between the three conditions (moral licensing, moral cleansing, and the neutral control condition) of Sachdeva et al.'s Study 1 and Study 3 were significant. For Study 1, no statistics on post-hoc comparisons were reported. When calculating the Cohen's d effect sizes comparing the moral licensing

sizes (n = 14 to 17 per condition), the obtained effects have large variances, implying that the true effect sizes could range from very small to very large.

There are no published direct replication attempts of the methodologies of Sachdeva et al.'s (2009) studies. Conway and Peetz (2012) conducted a study that was similar to Sachdeva et al.'s Study 1. However, this was not a direct replication because they adapted the procedure and added extra manipulations. We sought to replicate the studies by Sachdeva et al. to obtain additional insight in the complete moral self-regulation framework by testing for both moral licensing and moral cleansing effects contrasted to a neutral control condition.

We conducted high-powered replications of Sachdeva et al.'s (2009) Study 1 and Study 3 in Dutch student samples with 95% statistical power based on the effect size of the original studies. We did a third study with a U.S. sample via Amazon Mechanical Turk (MTurk) with 95% power based on the effect size that we obtained in our meta-analysis on self-licensing (d = 0.26[13]; Chapter 2). This study examined both dependent variables of original Study 1 and 3 in a counterbalanced order. For all studies, we report how we determined our sample sizes, all data exclusions, all manipulations, and all measures.

Study 4.1. Replication of Sachdeva et al.'s (2009) Study 1

Participants

Using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) we calculated that at least 63 participants were needed to achieve 95% power for the effect size of Sachdeva et al.'s

with the neutral control conditions, we found that for both studies, the confidence intervals included zero, indicating *marginally* significant moral licensing effects.

[13] Note that this effect size is based on a preliminary version of the meta-analysis reported in Chapter 2. This effect size is slightly smaller than the final effect size in Chapter 2, because the inclusion criteria in this preliminary version were slightly different (i.e., this version included studies on self-regulation and comparisons between licensing and control conditions that fell under our definition of licensing but were not necessarily expected to lead to licensing effect by the original authors, and did not include the latest papers that were published and the latest data that we received).

Study 1 (2009; N = 46). We planned to collect data for one full week, and our sample consisted of 106 undergraduate students who participated for course credit. One was removed because this participant indicated a willingness to donate 100 euro to charity, more than 32 standard deviations from the mean donation response. The remaining 105 participants (25 males, 78 females, 2 unknown, M_{age} = 19.58) included native Dutch students (83.3%), non-native Dutch students (9.5%), and foreign students (4.8%). Participants were randomly assigned to either the positive trait (n = 35), negative trait (n = 34), or neutral control condition (n = 36).

Materials and procedure

Participants completed the study as the first of a series of experiments behind separate desks in the lab. The experimenter in the lab was blind to condition. Prior to the experiment, participants provided their informed consent. The experimenter guided participants to their desks and instructed them to complete the paper-and-pencil questionnaire.

We obtained the original paper-and-pencil questionnaire from Sachdeva et al. (2009) and translated these materials into Dutch (for all materials see online supplements, Figure 4.1). The cover story indicated that the study was about handwriting styles. Depending on the assigned condition, participants were exposed to either nine positive trait words, nine negative trait words, or nine neutral words and were asked to copy each word four times and think about each word for 5-10 seconds. Next, participants were asked to write a short story about themselves including the words they just copied.

After this manipulation, participants responded to some filler items. Subsequently, the main dependent variable was presented. Participants read that the lab, in an effort to increase social responsibility, asked all participants whether they would like to contribute to a worthy cause. If they would like to do so, they could pledge to make a small donation to any good cause of their choice. They were told that they would be reminded of their choice at a later time via a confirmation e-mail from the experimenter. Participants could select to which cause(s) they would like to donate

(cancer research, animal rights, ending world hunger, environmental preservation, human rights, veteran's affairs, or other) and how much they would be willing to donate (from €0 up to €10 or another specified amount). Finally, participants completed seven self-presentation items from the Self-Monitoring scale (Lennox & Wolfe, 1984) and a set of demographic measures.

Known differences from original study

The only known difference between our replication and the original Study 1 and 2 of Sachdeva et al. (2009) was that we ran this study in a lab at a Dutch university, while the original study was conducted in a lab at a U.S. university. When participants were asked to write about the positive trait, neutral, or negative trait words, we used the exact instruction of the original Study 2, which explicitly stated that participants should use the nine given words to *write a story about themselves*. This was not done in Sachdeva et al.'s Study 1, although it was intended that participants would do so. As such, for this replication, we combined the best of Sachdeva et al's Study 1 (including a neutral control condition) and Study 2 (the manipulation with the clearest instruction).

Results

Following our confirmatory analysis plan, we conducted Sachdeva et al.'s (2009) analysis to test the effect of writing about one's own positive traits, negative traits, or neutral words on donation amount. Table 4.1 contains the mean responses per condition and statistical test. There were no significant differences between the moral identity conditions on donation amount[14]. The results of an additional regression model including gender, age, and ethnicity indicated that none of these factors significantly predicted donation amount (all *p*s ≥ .321). A reviewer suggested that self-monitoring might moderate the observed effects. It did not, $p = .086$. Analysis details for all studies are available in the online supplements (Figure 4.1).

[14] A nonparametric independent-samples Kruskal-Wallis test (which controls for the skewness of the data), also found a non-significant effect, $H(2) = 0.36$, $p = .837$.

Exploratory analysis.

When reading the recalled stories, we noticed that 55.7% of the participants violated the instructions by not writing about themselves or by using the words in a negating way (for instance, 'Alyssa is a generally friendly person with a caring and compassionate disposition' or 'I am neither a very caring nor compassionate individual'). When we only used a post hoc selection of those that wrote about their own positive traits (n = 28) and compared it to the neutral control condition, there was still no effect (p = .756).

Study 4.2. Replication of Sachdeva et al.'s (2009) Study 3

Participants

Using G*Power (Faul et al., 2009) we calculated that we should include at least 96 participants in our study to achieve 95% power for the effect size that Sachdeva et al. (2009) obtained in their Study 3 (the original used N = 46). We planned to collect data for one full week, and our sample consisted of 150 undergraduate students who participated for course credit (27 males, 122 females, 1 unknown, M_{age} = 20.34) and included native Dutch students (87.3%), non-native Dutch students (7.3%), and foreign students (4.7%). All participants were randomly assigned to either the positive trait condition (n = 49), the negative trait condition (n = 52), or the neutral control condition (n = 49).

Materials and procedure

Participants first provided informed consent, and then completed the study as the first of a series of experiments. The lab experimenter was blind to condition. The experimenter led participants to a separate cubicle and instructed them to complete the paper-and-pencil questionnaire.

The materials were the same as those in Study 4.1 except that the dependent variable was a hypothetical commons dilemma. In this commons dilemma, participants imagined a scenario in which they are the manager of a mid-sized industrial manufacturing plant. They read that all manufacturers reached an agreement to install

filters to eliminate toxic gasses and to run these filters 60% of the time. Running a filter was costly for the manufacturing plant, but would be beneficial to society. To measure cooperative behavior, participants were asked to indicate what percentage of time they would operate the filters, indicated on an 11-point scale from 0 (labeled 0%) to 10 (labeled 100%).

After the main dependent variable, participants explained their decision and completed three secondary measures; they estimated (1) the percentage of other managers who would not cooperate, on the same 11-point scale; (2) the amount of environmental damage expected when the filters would be run less than the agreed 60% on an 11-point scale from 0 (none) to 10 (a great amount); and (3) the likelihood of getting caught when operating the filters less than 60% of the time on an 11-point scale from 0 (impossible) to 10 (certain). Finally, participants completed the seven self-presentation items from the Self-Monitoring scale (Lennox & Wolfe, 1984) and a set of demographic measures.

Known differences from original study

The only known difference compared to the original study is that we ran this study in a lab at a Dutch university instead of a U.S. university.

Results

Following our confirmatory analysis plan, we conducted Sachdeva et al.'s (2009) analysis to test the effect writing about one's own (im)moral traits on cooperation (the amount of time participants were willing to run the filters). There were no significant differences between the conditions on cooperative behavior (Table 4.1)[15]. Furthermore, there were no effects on the secondary variables (Table 4.2). The results of an additional regression model including gender, age, and ethnicity indicated that none of these demographic variables predicted cooperative behavior (all *p*s ≥ .257). Self-monitoring did not moderate the observed effects ($p = .787$).

[15] A nonparametric independent-samples Kruskal-Wallis test (which controls for the not normally distributed data), also showed no effect, $H(2) = 2.87$, $p = .238$.

Exploratory analysis.

We noticed that 48.5% of the participants violated the recall instructions and did not write about their own traits or used the words in a negating way. When we only used a post hoc selection of those who actually wrote about their own positive traits (n = 42), there was still no difference on cooperative behavior between the positive trait stories about oneself and the neutral control condition (p = .197).

Study 4.3. Replication of Sachdeva et al.'s (2009) Study 1 and Study 3 with a general U.S. population sample on MTurk

Participants

Whereas in Study 4.1 and Study 4.2, we based our sample size on a power analysis using the original studies, for Study 4.3 we did so based on the effect size of self-licensing that we obtained in the preliminary data of our meta-analysis (Chapter 2). We calculated with G*Power (Faul et al., 2009) that we would need at least 918 participants in our study to achieve 95% power to find a self-licensing effect. The sample was recruited on MTurk. We included an instructional manipulation check to prevent inattentive participants from starting the study (see Oppenheimer, Meyvis, & Davidenko, 2009). Participants were asked to provide an answer to three neutral questions about stories and were explicitly instructed to answer 'five' on the first question, and 'seven' on the second and third question. Participants who did not follow these instructions (n = 160) could not participate in our study. Our final sample consisted of 940 [16] participants (449 males and 491 females, M_{age} = 33.41) who participated in exchange for $1.80. All participants were randomly assigned to the positive trait condition (n = 306), the negative trait condition (n = 308), or the neutral control condition (n = 326).

[16] We set the target higher than 918 to ensure a minimum of 918 valid participants after data exclusion.

Materials and procedure

Participants completed the study materials via the Qualtrics survey program. Participants could subscribe to participate in our study entitled 'writing style and several questions' if they had an MTurk approval rate that was higher than 95% and if they lived in the U.S.

After finishing writing the stories with the positive traits, negative traits, or neutral words, participants answered the filler questions and both dependent measures from Sachdeva et al.'s (2009) Study 4.1 (donation amount) and Study 4.3 (cooperative behavior) in a counterbalanced order. Subsequently, participants completed the self-presentation items from the Self-Monitoring scale and a set of demographic measures.

Known differences from original study

The study was conducted online. We made two slight changes to these materials to increase the credibility of the online study. First, for the cover story, we instructed participants that the study was about general writing styles instead of handwriting, as the latter would not be believable in an online study. Second, we changed the donation measure. We told participants that 10 of them would be randomly selected to win an additional $10 MTurk worker bonus. They were then asked that if they were one of the winners, would they be willing to donate a portion of this bonus to a cause of their choice from a list (cancer research, animal rights, ending world hunger, environmental preservation, human rights, veteran's affairs, or other). Participants selected a cause and indicated the amount they would donate ranging from $0 to $10.

Results

Donations.

Following our confirmatory analysis plan, we conducted Sachdeva et al.'s (2009) analyses to test the effect of writing about (im)moral traits on how much participants would want to donate to a good cause. We controlled for order effects by including the order in which the two dependent variables were presented as a separate independent variable in the model. Order did not affect the donation amount, $F(1, 934)$

= 0.78, p = .378, ηp^2 = .001, nor was there an interaction effect of order with the manipulation of what words participants wrote about, $F(2, 934) = 0.42$, p = .656, ηp^2 = .001.

As Table 4.1 shows, there was a main effect of moral identity condition on donation amount[17]. Post-hoc Tukey tests indicated that participants in the negative trait condition donated more money than participants in the positive trait condition (p = .044) and participants in the neutral control condition (p = .020). There was no difference in donation amount between participants in the positive trait condition and participants in the neutral control condition (p = .729). Thus, we did not find a moral licensing effect, but we did observe a moral cleansing effect - the recall of negative traits increased subsequent moral behavior. Self-monitoring did not moderate the observed effects.

Of the demographic variables gender, age, education level, family income, and ethnicity, only age significantly influenced donation amount (β = .11, $t(930) = 3.36$, p < .001). When we included age as a covariate to the effect of the manipulation on donation amount, the effect of the manipulation remained significant, $F(2, 932) = 3.15$, p = .043, ηp^2 = .007.

Cooperative behavior.

Next, we conducted Sachdeva et al.'s (2009) analyses to test the effect of moral identity condition on cooperation in a hypothetical commons dilemma. The order in which the dependent variables were presented did affect cooperative behavior ($F(1, 934) = 11.20$, p = .001, ηp^2 = .012), with participants who first completed the donation dependent variable displaying slightly more cooperative behavior (M = 6.47, SD = 1.77) than participants who first completed this cooperative behavior dependent variable (M = 6.11, SD =1.52). The interaction between moral identity and order was not significant, $F(2, 934) = 0.83$, p = .438, ηp^2 = .002. We do not know why this order effect exists, but for the current study it is mainly important that we control for this possible influence

[17] A nonparametric independent-samples Kruskal-Wallis test (which controls for the skewness of the data), found a similar effect, $H(2) = 5.85$, p = .054.

by adding it as a factor in the analyses. As Table 4.1 shows, there was no main effect of moral identity on cooperative behavior[18], nor on the secondary variables (see Table 4.2). Again, self-monitoring did not moderate the observed effects.

Of the demographic variables, only one of the ethnicity dummy variables significantly influenced cooperation (with African Americans cooperating less than others, $\beta = -.19$, $t(930) = -3.11$, $p = .002$). When including ethnicity as a covariate, there was still no effect of moral identity condition on cooperative behavior, $F(2, 933) = 0.81$, $p = .447$, $\eta p^2 = .002$.

Exploratory analyses.

We noticed that 43.6% of the participants violated the recall instructions and did not write about their own traits or used the words in a negating way. Using solely the coded stories about oneself in our analyses, there was a main effect of moral identity condition on donation amount ($p = .020$) with participants in the negative trait condition donating more money than participants in the positive trait condition ($p = .017$) and in the neutral control condition ($p = .009$). There was no main effect of moral identity condition on cooperative behavior ($p = .495$).

[18] A nonparametric independent-samples Kruskal-Wallis test (which controls for the not normally distributed data), also found a non-significant effect, $H(2) = 2.68$, $p = .713$.

Table 4.1

Means, standard deviations, sample sizes, and test statistics for dependent variables in all studies

	Positive trait	**Neutral trait**	**Negative trait**	***F***	***p***	**ηp²**
Dependent Variable	*M* (*SD*)	*M (SD)*	*M* (*SD*)			
Study 4.1, *N* = 105	35	36	34			
Donation amount[19]	2.89 (3.64)	2.78 (3.83)	2.35 (3.28)	0.21	.810	.004
Study 4.2, *N* = 150	49	49	52			
Cooperative behavior	6.29 (1.14)	5.88 (1.49)	5.84 (1.37)	1.70	.187	.023
Study 4.3, *N* = 940	306	326	308			
Donation amount	4.52 (2.91)	4.60 (3.11)	5.10 (3.18)	3.20	.041	.007
Cooperative behavior	6.27 (1.58)	6.39 (1.78)	6.21 (1.60)	0.78	.457	.002

Note. For donation amount the answers could range from 0 to 10 euro (or participants could indicate another amount). For cooperative behavior, the answers indicate how long participants would choose to do the costly cooperative act in the scenario (run filters), also on a range from 0 (*the least helpful*) to 10 (*the most helpful*). The statistical test of Study 4.3 is the main effect, controlling for order effects. Effect sizes and Confidence Intervals for the moral licensing and moral cleansing effects can be found in the Forest plots (Figure 4.2 and Figure 4.3).

[19] For the main dependent variable, participants were asked to indicate how much money they wanted to donate to a good cause. If they did not answer this question, we interpreted their response as 0 euro. When these participants were excluded from the analysis, the results did not differ, $F(2, 81) = 0.29$, $p = .747$, $\eta p^2 = .007$.

Table 4.2

Means, standard deviations, and test statistics for secondary measures in Study 4.2 and 4.3

	Positive trait	**Neutral trait**	**Negative trait**	***F***	***p***	**ηp²**
	M (*SD*)	*M (SD)*	*M* (*SD*)			
Study 4.2, *N* = 150	49	49	52			
Expected cooperative behavior of others	4.64 (2.56)	4.57 (2.55)	4.28 (2.74)	0.26	.770	.004
Estimated likelihood of getting caught	5.92 (1.80)	5.61 (1.85)	5.62 (1.75)	0.47	.624	.006
Negative consequences for the environment	7.22 (1.62)	6.90 (1.95)	6.62 (1.88)	1.41	.248	.019
Study 4.3, *N* = 940	306	326	308			
Expected cooperative behavior of others	4.72 (2.91)	4.89 (3.08)	4.51 (2.86)	1.30	.273	.003
Estimated likelihood of getting caught	6.91 (2.07)	7.09 (2.14)	7.05 (2.02)	0.63	.535	.001
Negative consequences for the environment	5.84 (2.04)	5.98 (2.11)	5.93 (1.99)	.33	.721	.001

Note. For expected cooperative behavior of others, answers could range from 0 (*the least cooperative*) to 10 (*the most cooperative*). For estimated likelihood of getting caught, answers could range from 1 (*impossible*) to 10 (*certain*). For negative consequences for the environment, answers could range from 1 (*none*) to 10 (*a great amount*).

General discussion

We made three attempts to replicate the findings of Sachdeva et al. (2009) on moral licensing, with samples based on pre-calculated power and pre-planned analyses. In the first two replication attempts using student samples, the data did not confirm the original results. In our third replication attempt using a general population sample the data did not confirm the moral licensing effect. We did, however, find support for the moral cleansing effect on one of the two dependent variables in Study 4.3, but not in Study 4.1 and 4.2.

Current status of the moral licensing effect

We conducted a meta-analysis of this moral licensing effect by including both the original Study 1 and 3 by Sachdeva et al. and the three current replication attempts, using the metafor package of Viechtbauer (2010). For our Study 4.3, we used the average effect size of the two dependent variables. The random effects meta-analysis including all five studies produced a mean effect size of moral licensing of d = 0.07 ([CL_{95}] -0.20 to 0.35). There was thus no significant moral licensing effect across studies (z = 0.52, p = .603). Figure 4.2 contains an overview of all moral licensing effect sizes (when compared to the neutral control conditions).

Current status of the moral cleansing effect

We conducted a meta-analysis of the moral cleansing effect by including both the original Study 1 and 3 by Sachdeva et al. and the three current replication attempts. The random effects meta-analysis including all five studies produced a mean effect size of moral cleansing of d = 0.04 ([CL_{95}] -0.11 to 0.20). There was thus no significant moral cleansing effect across studies (z = 0.53, p = .593). Figure 4.3 contains an overview of all moral cleansing effect sizes (when compared to the neutral control conditions).

Possible limitations of our replication attempts

Although we did our best to design direct replications of the original studies, differences are inevitable, and some of those may be consequential for moderating the results. First, our Study 4.1 and 4.2 used Dutch students not U.S. students. There is no theoretical reason to expect different licensing effects for Dutch compared to U.S. citizens, but our pilot test (see online supplements) suggested that words in the positive moral trait condition were seen to be slightly more positive in the U.S. than in the Netherlands. Even so, the words were evaluated very positively in both national samples. Study 4.3 used a U.S. based sample, but this study differed on two aspects compared to the original study. It was conducted online instead of in the lab, and the manipulation involved donating a part of potential winnings instead of money out-of-pocket. We cannot rule out that these procedural differences were consequential, but

there presently exists no theoretical reason or identification of these as boundary conditions on moral licensing.

Conclusion

Although Sachdeva et al. (2009) theorized that moral licensing and moral cleansing should be considered jointly as being part of a moral self-regulation process, our three high-powered studies did not replicate the key moral licensing effect. Further, the meta-analytic result suggests that the present state of evidence with this paradigm is not different from a null effect. Sachdeva et al. (2009, p. 524) suggested that their findings showed that "moral-licensing and moral-cleansing effects can act convergently as part of a moral self-regulation process". Based on the present findings, we do not argue that the theory is incorrect, only that it lacks sufficient empirical support when using the Sachdeva et al. (2009) paradigm.

We suggest three concrete steps to clarify the effects of moral licensing on social judgment. First, the method used by Sachdeva et al. (2009) seems unlikely to elicit moral licensing, especially since many participants violated the recall instructions and did not write about their own traits or used the words in a negating way. This is a procedure specific issue; it does not invalidate moral licensing more generally. Second, the meta-analysis in Chapter 2 suggests that the effect is relatively small. Therefore, small sample studies are highly inadvisable as they would need to leverage chance to detect a result using null hypothesis significance testing. Third, because moral licensing and moral cleansing are theoretically distinct, it is important to use a neutral control condition to clarify the role of each in social judgment.

The study reported in this article
earned *Open Data*, *Open Materials*,
and *Preregistered* badges:
https://opensciencefframework.org/project/3cmz4/

Figure 4.1. Access to the data and materials of Chapter 4.

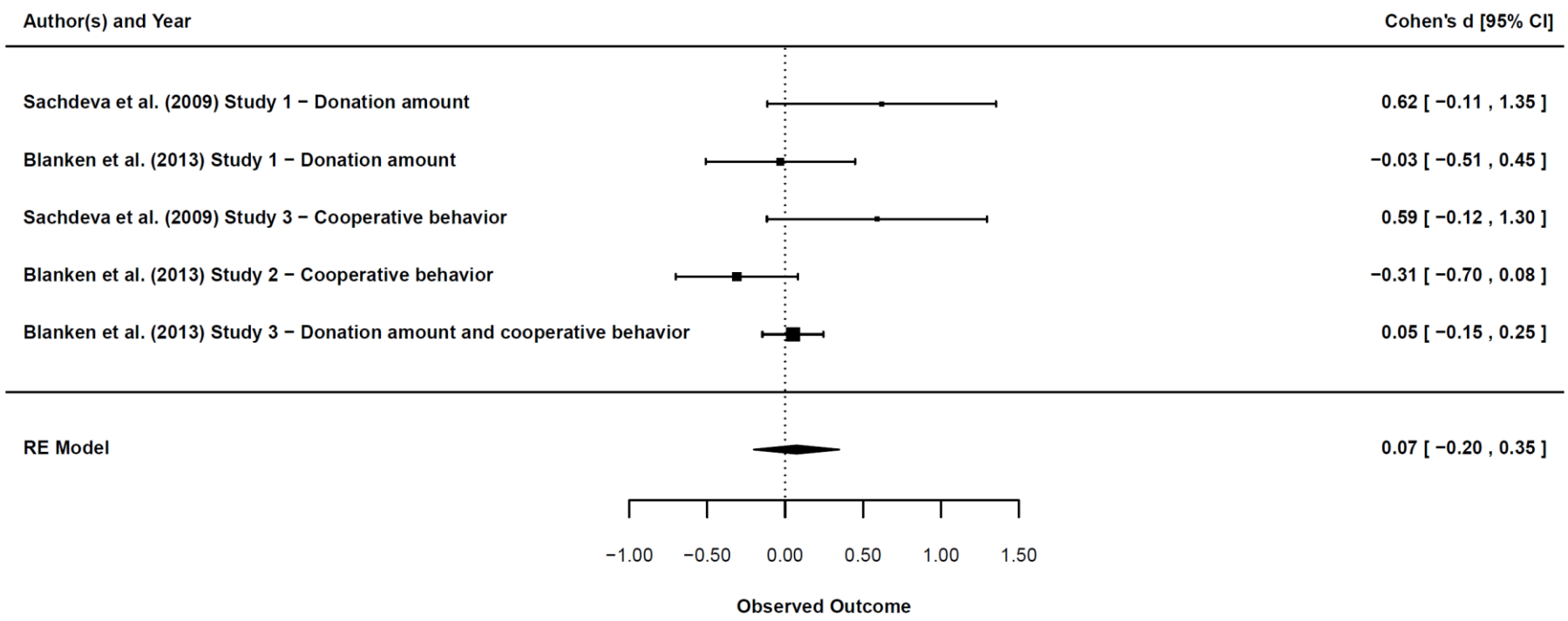


Figure 4.2. Forest plot including all comparisons between the *moral licensing* and neutral control conditions of the original studies by Sachdeva et al. (2009) and our replication attempts.

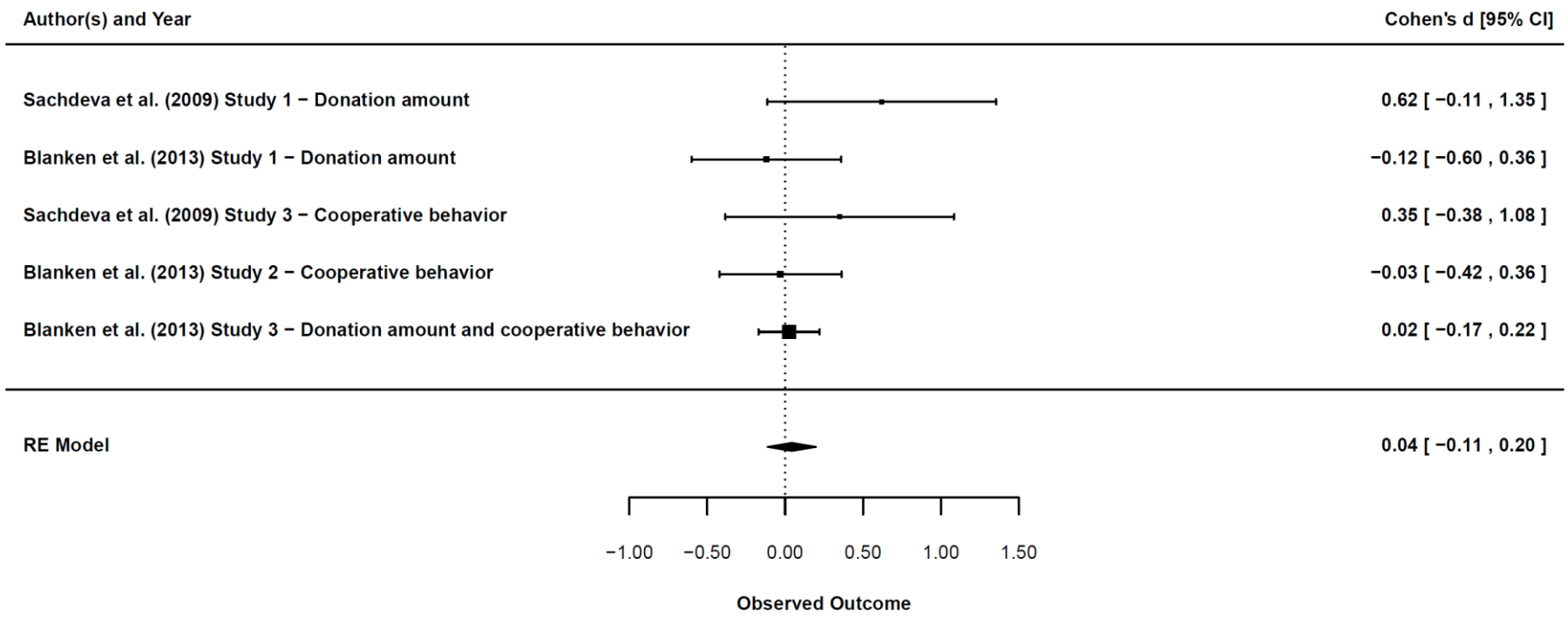


Figure 4.3. Forest plot including all comparisons between the *moral cleansing* and neutral control conditions of the original studies by Sachdeva et al. (2009) and our replication attempts.

Part II: Novel Perspectives on Self-Licensing[20]

[20] Note that the first part of this dissertation predominantly focused on licensing in the moral domain (i.e., moral licensing), whereas the second part mostly focuses on licensing related to goal-directed self-regulatory behaviors. As outlined in Chapter 1, self-licensing can lead to a broad spectrum of undesirable behaviors, both at the individual and societal level. For both the social and the individual domain, the underlying processes of the licensing effect seem very similar. Importantly, research on moral self-regulation (Zhong, Liljenquist, & Cain, 2009) shows that similar to other self-regulatory behaviors, moral behavior is characterized by internal regulations where the desired 'moral self' motivates goal achievement, and not living up to moral aspirations leads to emotional distress (Higgins; 1996; 1987).

CHAPTER 5

Good Deeds and Temptations:

Two Ways in Which Self-Licensing can be Triggered

Self-licensing theory posits that people who previously behaved in a good way more easily feel justified to subsequently engage in undesirable behavior. In the current chapter we investigated whether there are two different ways in which self-licensing can be triggered. We proposed that self-licensing can not only occur when a prior good deed makes one more likely to engage in subsequent questionable behavior ('good deed licensing'), but it can also be a justification strategy that one can deliberately use when tempted to engage in undesirable behaviors ('temptation-based licensing'). When self-categorizing (Study 5.1) or manipulating (Study 5.2) these two ways of licensing, the large majority of the participants indicated to regularly experience both good deed and temptation-based self-licensing. Importantly, we found that both ways of licensing have different antecedents and affective responses, highlighting the importance for researchers in the field to distinguish between these two ways self-licensing.

This chapter is based on: Blanken, I., Van de Ven, N., & Zeelenberg, M. (2015b). Good deeds and temptations: Two ways in which self-licensing can be triggered. *Manuscript under review*.

Self-licensing theory posits that people who behave in a good way, later feel justified to perform bad behaviors (Miller & Effron, 2010; for recent overviews, see Chapter 2; Effron & Conway, 2015). For example, someone who has just worked out at the gym might later feel justified to eat a large piece of chocolate cake. In a similar vein, a CEO of a large company who decided to invest a lot of money in an environmental sustainability program might later find it more acceptable to file fraudulent tax returns. Self-licensing is related to a variety of individual and social behaviors in various domains such as racism (Choi, Crandall, & La, 2014; Effron, Cameron, & Monin, 2009; Effron, Monin, & Miller, 2012; Mann, & Kawakami, 2012), job hiring (Cascio & Plant, 2015; Monin & Miller, 2001), sustainability (Mazar & Zhong, 2010), consumerism (Khan & Dhar, 2006), ethics (Conway & Peetz, 2012; Jordan, Mullen, & Murnighan, 2011), and self-regulation (Chiou, Yang, & Wan, 2011; De Witt-Huberts, Evers, & De Ridder, 2012; Mukhopadhyay, Sengupta, & Ramanathan, 2008; Mukhopadhyay & Venkataramani, 2009). Self-licensing seems closely related to theories on entitlement, which posit that people can feel 'entitled' to take more than others because of their previous positions and behaviors (Campbell, Bonacci, Shelton, Exline, & Bushman, 2004). As self-licensing has adverse consequences for such a wide range of behaviors, research on this topic can help explain why people feel entitled to and actually engage in bad behavior.

A close look at the definitions of self-licensing lends support to the idea that self-licensing is typically perceived as a process in which the initial good act later influences one's behavior: "Self-licensing occurs when past moral behavior makes people more likely to do potentially immoral things without worrying about feeling or appearing immoral" (Merritt, Effron, & Monin, 2010, p. 344); "Virtuous acts can license subsequent asocial and unethical behaviors" (Mazar & Zhong, p. 495.); "A prior intent to be virtuous boosts people's self-concepts, thus reducing negative self-attributions" (Khan & Dhar, 2006, p. 259). Miller and Effron (2010) argued that prior good behavior can create either a *credit*, which can later be used to trade in to transgress, or can serve as a *credential*, where the good deed creates a virtuous reputation after which a transgression is seen as less bad. Thus, definitions of self-licensing clearly suggest that a prior good deed can lead to undesirable behavior.

The way self-licensing is typically perceived reflects how self-licensing is usually investigated in the lab: Initial good behavior is manipulated (and contrasted with for example, a neutral condition) and the effect of this manipulation on less desirable behavior is measured. For instance, participants who established their non-racist attitudes by voting for president Obama (Effron et al., 2009) or through preferring a black person for a consulting firm job (Monin & Miller, 2001) were subsequently more likely to make prejudiced judgments. Furthermore, participants who purchased green products in a virtual shopping paradigm offered less money in a subsequent ultimatum game and stole more money compared to participants who purchased regular products (Mazar & Zhong, 2010).

Thus, research on self-licensing focuses on a specific sequence of two consecutive behaviors or events, where initial good behavior leads to less desirable behavior. This sequence is also reflected in how self-licensing is typically investigated. In the current chapter we investigate whether the reverse pattern also occurs, in which the temptation to display undesirable behavior leads to a retrospective search for a license.

We believe that there are two different ways in which self-licensing can be triggered. When people recalled self-licensing examples from their daily lives in informal discussions, we noticed that some people hightlighted the intital good behavior as the initiator of the licensing effect (e.g., "I did yardwork and therefore I can eat pizza"), while others emphasized the temptingness of displaying the undesirable behavior as the main cause of licensing (e.g., "I really wanted to eat pizza and I can because of all the yardwork I just did"). Although these two expressions might seem minor variations of the same thing, we think they represent very different ways in which licensing may come about. It seems that licensing effects not only arise after people engaged in good behaviors (as is typically assumed), but also when they are tempted to engage in certain behaviors and are therefore motivated to search for reasons that justify this.

The latter reasoning fits with a justification-based account of self-regulation (De Witt-Huberts, Evers, & De Ridder, 2014a), which implies that people who face a dilemma in which they would like to display undesirable or bad behaviors develop and employ

justifications that allow performing these behaviors. De Witt-huberts, Evers, and De Ridder (2014b) found a correlation between the self-rated temptingness of indulgent behavior and the number of justifications that people indicate or generate to justify giving in to this temptation. Specifically, the more tempting people thought a high caloric chocolate bar was, the more reasons they wrote down why it was acceptable to eat it. Moreover, in Chapter 6 we show that when people are exposed to a temptation, prior good deeds become seen as a reason to justify giving in to the temptation. These findings suggest that licensing can also take place in a reverse order, where the tempting bad behavior leads to a search for a license. This seems in line with theorizing on motivated reasoning (Kunda, 1990), which posits that when people have a preferred conclusion, they are likely to construct justifications that are supportive of that particular conclusion. So, when people are tempted to perform a bad deed, people may construct justifications by mentally scrolling through their prior actions until they find good or moral behavior that helps them justify performing the bad deed.

We believe that investigating the different ways in which licensing may come about can serve as an initial step to establish a robust paradigm that documents the licensing effect. Despite a substantial number of published licensing studies, it is not always easy to replicate the 'traditional' licensing effect, where good behavior leads to less desirable behavior (cf. Chapter 4). A meta-analysis including 91 experimental studies (Chapter 2) revealed that the average effect size of moral licensing (i.e., self-licensing in the moral domain) is small-to-medium (Cohen's d = 0.31). The meta-analysis could not confirm any of the moderators that were theorized to be of importance. Moreover, a p-curve analysis (Simonsohn, Nelson, & Simmons, 2014a,b) indicated that research on moral licensing lacks evidential value (Chapter 3). Together, these findings call into question the empirical support for the licensing effect. It therefore becomes a question again whether licensing exists, and if so, how it functions.

The current chapter

There is doubt concerning the strength of the support for (moral) licensing. This implies that researchers would do well to restart thinking the paradigm with which licensing is being studied. To do so, we think it is well to adopt a full cycle approach

(Cialdini, 1980) of self-licensing. Mortensen and Cialdini (2010, p.53) argue that when studying a novel phenomenon researchers should first

> use naturalistic observation to determine an effect's presence in the real world, theory to determine what processes underlie the effect, experimentation to verify the effect and its underlying processes, and a return to the natural environment to corroborate the experimental findings.

We think the initial step, testing whether and how people experience licensing in their daily lives, has not extensively been made yet in the self-licensing literature[21]. If we need to rethink the paradigms we use to study licensing, starting with examining how it exists in real life is the valuable first step to start with.

Besides starting the full-cycle approach with natural observations, we propose that there are two different ways in which self-licensing can come about: A good deed can permit one to display undesirable behavior (*good deed self-licensing)*, or the temptation to display undesirable behavior can initiate a search for a license (*temptation-based self-licensing)*. We will explore to what extent both ways of self-licensing occur in daily life. In addition, we will investigate the antecedents and consequences of both ways. Because we propose that initial good behaviors play an important role in the emergence of good deed self-licensing, and the temptingness of the undesirable behavior is an essential trigger of temptation-based self-licensing, we will explore the effects of the 'goodness' of the prior good behavior, the 'badness' and temptingness of the undesirable behavior. To draw a completer picture of both ways of licensing, we will also explore the affective consequences.

[21] A notable exception to the lack of naturalistic observations of self-licensing is the work of Hofmann, Wineski, Brandt, and Skitka (2014), who made a first step in investigating moral licensing in daily life. In a general study on moral behavior in daily life, participants (N = 1252) indicated five times per day whether they commited, were the target of, witnessed, or learned about a moral or immoral act within the past hour. They studied several research questions on morality in everyday life, and the licensing effect was one of them. Hofmann et al. found a moral licensing pattern, in that people who committed a moral act had a larger likelihood of committing an immoral act later that day.

Study 5.1 explores the self-licensing effects people recall from their daily lives and investigates to what extent these effects reflect good deed self-licensing versus temptation-based self-licensing. In Study 5.2 participants are randomly assigned to recall either good deed licensing or temptation-based licensing and we explore the different antecedents and affective consequences. These studies will serve as a first step to verify different manifestations of the self-licensing effect in daily life. This will enable researchers in the field to more effectively comprehend what self-licensing is, how it works, and when it occurs.

Study 5.1. Recalls of self-licensing

In Study 5.1, we explore to what extent self-licensing effects occur in daily life and whether these effects are triggered by good deeds (good deed self-licensing) or the desire to justify displaying undesirable behaviors (temptation-based self-licensing). We explained undergraduate psychology students the concept of self-licensing. Subsequently, they indicated whether they themselves ever experienced it, and if so they were asked to recall their latest experience of self-licensing and answer several questions about this situation.

Method

Two hundred and seventy-five Tilburg University undergraduate psychology students (72 males and 203 females, M_{age} = 20.16) participated in a 60-minute research session with several unrelated experiments. At the start of our study, participants read the following:

> *In the past few years, psychologists investigated the 'self-licensing effect', which implies that people allow themselves to engage in certain behaviors because they previously displayed good behaviors.*

Subsequently, participants were asked whether this ever occurred to them. Participants who indicated that this had ever occurred to them were asked to recall the most recent situation in which they allowed themselves to display certain behaviors after previous good behavior. To obtain more insight in the natural occurrence of self-

licensing, participants were then asked to indicate when this situation took place (see Table 5.1 for answer options) and whether others were present during these behaviors (yes / no). They were also asked to specify how they felt after displaying these behaviors (feeling in general, -3 = *not good at all*, 3 = *very good*; happiness, satisfaction, regret, guilt, 0 = *not at all*, 6 = *very much*).

To test our proposition that licensing can be triggered both by good deeds and the desire to justify displaying undesirable behaviors, participants were asked to indicate which of the following descriptions corresponded best to the licensing episode they had recalled: a) after displaying good behavior, I felt that I had permission to display certain behaviors; b) when I wanted to display certain behaviors, I justified these behaviors through my previous good behaviors; c) I do not know. Finally, participants indicated to what extent they regarded the licensed behavior that they engaged in as a reward for their prior good behavior (1 = *not at all*, 7 = *very much*). The domains in which the good behaviors and licensed behaviors took place were coded by the first author (these data can be found in Appendix 5.1).

Results and discussion

Almost all participants indicated that they experienced self-licensing effects (95.3%, *n* = 262)[22]. The large majority of the participants experienced such an effect within the last month (Table 5.1). In addition, most participants (67.3%) indicated that the licensed behaviors took place in the same domain as the previous good behaviors. For example, the prior good deed was 'working very hard', after which the participant allowed himself a free evening in which he did nothing but relax. Participants displayed the licensed behaviors both in the presence of others (56.5%) and alone (43.5%).

Most participants indicated that their recalled situation resembled good deed self-licensing, where prior good behavior leads to subsequent bad behavior (53.8%). Also, a substantial number of participants (37.0%) indicated that their recalled situation

[22] Participants who indicated that they did not experience self-licensing effects (*n* = 13) did not fill out the other questions.

resembled temptation-based self-licensing, where displaying bad behavior was justified using the prior good behavior. This points to our initial idea that many of the cases of licensing might actually start with a temptation to do something with more or less undesirable consequences, for which the prior good deed is used as a justification. The rest of the participants (9.2%) indicated that they did not know which of the two descriptions best resembled their recalled behaviors.

Table 5.1

Timing of most recent recalled self-licensing experience in Study 5.1

Timing	**Percentage of respondents**	***n***
Within the past seven days	42.7%	112
Within the past month	46.6%	122
Within the last two months	5.3%	14
Within the last year	5.3%	14

Note that we do not wish to make the claim that these percentages of the two suggested pathways to licensing are hard and exact numbers. First, in the instructions we had to choose how to describe the licensing process. We chose to describe it as 'doing something *because* you did a good deed before'. Both ways of licensing fit this described process, but it seems better suited to trigger examples of good deed licensing (as it interprets the good deed as the trigger of licensing) than temptation-based licensing (that interprets the temptation as the trigger). If we had chosen to describe licensing as "wanting to do something that was undesirable, but allowing yourself to do it anyway because of something good you have done before", we would have likely found relatively more examples of temptation-based licensing. Similarly, perhaps temptation-based licensing is a less socially desirable response, because admitting that one aimed to justify doing something undesirable might be seen as less positive than indicating that one felt licensed to display bad behavior after displaying good

behavior. The exact numbers should thus be interpreted with care, but the fact that a large proportion of participants indicated that their recalled self-licensing episode fitted better with temptation-based licensing seems an important novel finding.

Interestingly, these two ways of licensing differently affected how people felt after displaying licensed behaviors (Table 5.2). Participants who indicated that their recalled situation resembled good deed licensing experienced more positive and less negative emotions after displaying the self-licensed behaviors, compared to participants who indicated that their behavior resembled temptation-based self-licensing. In addition, participants who indicated that their recalled situation resembled good deed licensing regarded their licensed behavior more as a reward for their good behavior compared to participants who indicated that their behavior resembled temptation-based licensing.

To conclude, the results of Study 5.1 show that people can quite easily recall self-licensing experiences from their daily lives. In fact, 95.3% of the participants could recall a licensing episode and 89.3% of these participants indicated that the latest experience of it had occurred in the last month. This frequency of occurrence testifies to the importance of self-licensing theory. This frequency-estimate is even likely to be an underestimation, as these are only the examples people were aware of while some licensing perhaps occurs without conscious awareness (e.g., Khan & Dhar, 2006). Interestingly, these results show that recalls of self-licensing experiences can be differentiated in two sizeable categories of responses: good deed licensing or temptation-based licensing. The different ways of self-licensing also seem to lead to different emotional effects. In Study 5.2, we manipulate these different ways of self-licensing to further explore the different antecedents and consequences of both ways.

Table 5.2

Affective responses after engaging in self-licensing of participants separated by self-classified good deed and temptation-based ways of self-licensing in Study 5.1

	Good deed licensing *n* = 141	**Temptation-based licensing** *n* = 97			
Emotional responses after licensed behavior	*M (SD)*	*M (SD)*	*F*(1,236)	*p*	η_{p^2}
Overall valence	5.80 (1.12)	4.78 (1.42)	83.30	<.001	.14
Happiness	4.47 (1.13)	3.62 (1.42)	26.45	<.001	.10
Satisfaction	4.56 (1.12)	3.51 (1.49)	38.82	<.001	.14
Guilt	1.30 (1.34)	2.58 (1.61)	44.13	<.001	.16
Regret	1.17 (1.63)	2.49 (1.62)	45.06	<.001	.16
Reward	4.71 (1.51)	5.73 (1.03)	38.31	<.001	.14

Note. Participants indicated their feeling in general from -3 (*not good at all*) to 3 (*very good*), the extent to which they experienced different emotions from 0 (*not at all*) to 6 (*very much*), and the extent to which they experienced the licensed behavior as a reward for their good behavior from 0 (*not at all*) to 7 (*very much*).

Study 5.2. Recalling the different ways of licensing

In Study 5.2, we manipulated the recall of the different ways of licensing by explaining participants about self-licensing either as initiated by the good deed (Good Deed Licensing condition) or as initiated by the desire to display undesirable behavior (Temptation-Based Licensing condition). Similar to Study 5.1, participants indicated whether they ever experienced such effects, recalled their latest experience of this effect, and answered several questions about this situation. Furthermore, we now used a sample recruited via Amazon Mechanical Turk (MTurk), that yields a sample that is broader in age range and education level than the sample we had used for Study 5.1.

Method

Five hundred and one U.S. based participants (308 males and 193 females, M_{age} = 31.79) completed our study on MTurk. Participants were randomly assigned to the Good Deed Licensing condition (n = 260) or to the Temptation-Based Licensing condition (n = 241). Participants in the Good Deed Licensing condition read the following:

> *Research has shown that after doing something good, people often allow themselves to do something that they would normally refrain from (e.g., I did good deed A, therefore I feel I can now allow myself to do B). Do you recognize this in your own behavior?*

Conversely, participants in the Temptation-Based Licensing condition read:

> *Research has shown that when people want to do something they would normally refrain from, they often justify these behaviors by thinking about something good they did before (e.g., I wanted to do A, and allowed myself to do so because I had done good deed B before). Do you recognize this in your own behavior?*

Participants were asked whether this ever occurred to them. Participants who indicated that this had ever occurred to them were asked to recall the most recent situation where they recognized the described behaviors. Next, they indicated when this situation took place, how much time there was between the good behavior and licensed behavior, and whether other people were present during this situation. Subsequently, participants evaluated both their good and their licensed behaviors (both behaviors were rated on three dimensions: -5 = *very bad*, 5 = *very good;* -5 = *very negative*, 5 = *very positive;* -5 = *very unfavorable*, 5 = *very favorable,* $\alpha_{\text{good behavior}}$ = .94, $\alpha_{\text{licensed behavior}}$ = .93). In addition, participants were asked to indicate how tempting and how attractive the licensed behavior was (1 = *not at all*, 7 = *very much,* r = .50, p < .001) and how they felt after displaying these behaviors (overall valence, -5 = *not good at all*, 5 = *very good*; happiness, satisfaction, regret, guilt, 0 = *not at all*, 6 = *very much*). Finally, participants indicated to what extent they regarded the licensed behavior that they displayed as a reward for their prior good behavior (1 = *not at all*, 7 = *very much*). The

domains in which the good behaviors and licensed behaviors took place were coded by the first author (these data can be found in Appendix 5.1).

Results and discussion

The majority of the participants indicated that they experienced self-licensing effects; this was the case in both the Good Deed Licensing condition (76.9%) and the Temptation-Based Licensing condition (73.9%)[23]. The percentage of participants that recognized the specified self-licensing in themselves did not differ between conditions, χ^2 (1, N = 501) = 0.63, p = .426. Note that the percentage of participants who indicated that they did not experience these effects in daily life is higher than in Study 5.1, where it was only 4.7% of the participants who did not experience such effects. Specifying the recall of self-licensing to one of the two ways of self-licensing thus likely reduced the chance that people recognized it in their own behavior. Another possibility would be that this sample is more representative than our student sample in Study 5.1. Still, a substantial amount of people experienced self-licensing and also did so within the last month, confirming that self-licensing occurs to most people and rather frequently as well.

A nonparametric Mann-Whitney U test (as responses were given on an ordinal scale, see Table 5.3) showed that participants in the Good Deed Licensing condition indicated that the moments of recent self-licensing experiences took place (marginally) further in the past than participants in the Temptation-Based Licensing condition, U = 15991, p = .087. In addition, participants in the Temptation-Based Licensing condition indicated more time had passed between their previous good behavior and their licensed behaviors, compared to participants in the Good Deed Licensing condition, U = 12937, p < .001 (see Table 5.3 for responses).

Participants displayed the licensed behaviors both in the presence of others (57.7%) and alone (42.3%), with no difference between conditions, χ^2(1, N = 378) = 0.020, p = .917. Table 5.4 contains the responses to the questions on how people evaluated both

[23] Participants who indicated that they did not experience self-licensing effects did not fill out the other questions.

the prior good behavior and the licensed behavior (how bad or good the behavior was seen to be), with the tests on whether the responses differed between conditions. Participants in the Temptation-Based Licensing condition indicated that the licensed behavior they displayed appeared to be more negative than participants in the Good Deed Licensing Condition (this difference was marginally significant). In addition, participants in the Temptation-Based Licensing condition indicated that the licensed behavior was more tempting and more attractive to them than participants in the Good Deed Licensing condition. There was no difference in the evaluation of previous good behaviors between both conditions. This shows that for temptation-based licensing, the licensed behavior was both more tempting and attractive, but at the same time also appeared to be somewhat more negative. In other words, this was a stronger temptation with both more desirable and undesirable characteristics. Finally, in Study 5.1 we found that participants who self-classified their licensing episode as good deed self-licensing more often saw their licensed behavior as a reward for their prior good deed, but this was not the case in this study. We have no clear explanation why this differed from Study 5.1.

Table 5.3

Recency of latest recalled self-licensing experience and time between the good and the licensed behaviors in Study 5.2

	Good Deed Licensing *n* = 200	**Temptation-Based Licensing** *n* = 178
Moment of last episode		
Within the past seven days	117 (58.0%)	115 (64.6%)
Within the past month	40 (20.0%)	27 (20.8%)
Within the last two months	16 (8.0%)	12 (6.7%)
Within the last year	23 (11.5%)	14 (7.9%)
Earlier	4 (2.0%)	0 (0.0%)
Time between the good and the licensed behaviors		
+/- one hour	97 (48.5%)	45 (25.3%)
+/- one day	54 (27.0%)	58 (32.6%)
+/- one week	35 (17.5%)	47 (26.4%)
+/- one month	5 (2.5%)	15 (8.4%)
> one month	9 (4.5%)	13 (7.3%)

Table 5.4

Evaluations of behavior for the Good Deed Licensing and the Temptation-Based Licensing conditions in Study 5.2

	Good Deed Licensing n = 200	**Temptation-Based Licensing** n = 178			
	M (SD)	*M (SD)*	*F*	*p*	η_{p^2}
How good / positive / favorable would you rate the good behavior?	3.59 (1.53)	3.34 (1.75)	2.12	.146	.01
How good / positive / favorable would you rate the licensed behavior?	-0.48 (2.49)	-0.96 (2.42)	3.53	.061	.01
To what extent do you regard the licensed behavior as a reward for your previous good behavior?	5.35 (1.52)	5.34 (1.29)	0.00	.987	.00
How tempting / attractive was the licensed behavior?	5.46 (1.26)	5.74 (1.00)	5.61	.018	.02

Note. Participants evaluated both their good and their allowed behaviors on three dimensions (-5 = *very bad*, 5 = *very good;* -5 = *very negative*, 5 = *very positive;* -5 = *very unfavorable*, 5 = *very favorable).* For both the good and the allowed behavior, the answers on these three items were averaged into one scale (α = .94 and .93, respectively). Participants rated the rewarding value of their good behavior on one item: 'to what extent do you regard the behavior that you allowed yourself to display as a reward for your previous good behavior?' (1 = *not at all*, 7 = *very much).* Participants rated the temptingness of their allowed behaviors on two items: 'how tempting was this behavior for you?' and 'how attractive was this behavior for you' (1 = *not at all*, 7 = *very much*) (*r* = .50). These items were averaged into one item.

Table 5.5 displays how positive and negative people felt after displaying the licensed behavior, with the statistical tests comparing the differences between conditions. There was no difference between conditions in how good (or bad) people felt themselves in general. For specific positive emotions, there was no effect for happiness and only a marginally significant effect for satisfaction (in that participants in the Good Deed

Licensing condition felt more satisfied than those in the Temptation-Based Licensing condition). This is different from Study 5.1, where participants who self-classified their licensing episode as good deed licensing felt better, happier and more satisfied than participants who classified it as temptation-based licensing. For guilt and regret we find the same as we did in Study 5.1: Participants who recalled an episode of temptation-based licensing felt more regret and guilt than participants who recalled an episode of good deed licensing.

Table 5.5

Emotional responses for the Good Deed Licensing and the Temptation-Based Licensing conditions in Study 5.2

	Good Deed Licensing, *n* = 200	**Temptation-Based Licensing,** *n* = 178			
Emotional responses	*M (SD)*	*M (SD)*	*F*	*p*	η_{p^2}
Overall valence	0.85 (2.59)	0.61 (2.67)	0.74	.390	.00
Happiness	3.65 (1.74)	3.42 (1.75)	1.70	.193	.01
Satisfaction	4.12 (1.71)	3.81 (1.69)	3.15	.077	.01
Guilt	2.33 (1.83)	2.78 (1.86)	5.53	.019	.01
Regret	2.27 (1.85)	2.80 (1.91)	7.77	.006	.02

Note. Participants evaluated their general feeling from -5 = *very bad* to 5 = *very good.* All specific emotions were rated from 0 = *not at all* to 6 = *very much.*

To conclude, the results of Study 5.2 show that self-licensing can come about in two different ways. The majority of participants was able to recall self-licensing effects in both the Good Deed Licensing condition (76.9%) and the Temptation-Based Licensing condition (73.9%). Furthermore, the results show that temptation-based licensing is more likely to arise when the behavior one would normally refrain from is more tempting (and perhaps more negative, this latter result being marginally significant). Likely as a consequence of this, people experience more negative emotions (guilt,

regret) after temptation-based licensing compared to good deed licensing. In support of this reasoning, there was a clear correlation between the negativity of the behavior that one would normally refrain from and the negative emotions after displaying this behavior (r = .37, p < .01). Our results also reveal that people who experience temptation-based licensing seem to go further back in time to find a good reason as there is more time between the prior good behavior and the licensed behavior than for people who recalled an episode of good deed licensing. Participants who experienced good deed self-licensing did not regard the good behavior as more good, positive, or favorable than participants who experienced temptation-based self-licensing. Thus, whereas temptation-based licensing is characterized by more tempting and more undesirable behavior one would otherwise refrain from, good deed licensing is not characterized by more positive good behaviors. Combined with the results of Study 5.1, these findings support our theorizing that licensing can be initiated through both good behavior and the temptation to engage in less desirable behavior, and that both ways of licensing lead to different affective responses.

General Discussion

The purpose of the present chapter was to investigate whether there are two different ways in which self-licensing can be triggered. We proposed that self-licensing can not only occur when a prior good deed makes one more likely to engage in subsequent questionable behavior (good deed licensing), but it can also be a justification strategy that one can deliberately use when tempted to engage in undesirable behaviors (temptation-based licensing). To test this proposition, we explored to what extent both ways of self-licensing occur in daily life. Our studies show that the large majority of people experiences self-licensing in their daily lives. People recalled the most recent episode they could think of; in more than 80% of the cases the licensing episode occurred within the last month. These results testify to the importance of self-licensing, as most people experience it regularly. Importantly, our findings show that people experience both good deed and temptation-based licensing.

In support of this distinction between two possible ways of self-licensing, participants categorized their own recalled episodes of licensing in Study 5.1. The results confirmed

that most participants categorized their licensing experiences as good deed licensing, but also a considerable percentage of the participants indicated to have experienced temptation-based licensing. Furthermore, testifying to the qualitative different experiences of both ways of self-licensing, those who classified their experiences as temptation-based licensing experienced fewer positive emotions and more negative emotions after displaying the licensed behavior than participants who self-classified their experiences as good deed licensing. In Study 5.2, where we specifically asked participants to either recall a temptation-based licensing episode or a good deed licensing episode, the results showed that temptation-based licensing is more likely to be activated when the behavior one would normally refrain from is both more tempting *and* marginally more negative. Again, participants experienced more negative emotions after temptation-based licensing experiences compared to good deed licensing experiences. Together, these findings support our theorizing that both ways of self-licensing exist, and that both ways lead to different affective responses

Participants who experienced temptation-based self-licensing indicated that there was more time between the initial good behavior and the licensed behavior compared to participants who experienced good deed self-licensing. Together with the findings that the licensed behaviors of the temptation-based way of licensing were classified as more tempting and marginally more negative, this suggests that people are 'willing' to go further back in time to find a reason to engage in more tempting and more desirable behavior. This seems consistent with motivated reasoning (Kunda, 1990), which posits that when individuals want to draw a conclusion they search for justifications to support this particular conclusion. Thus, the temptation to display undesirable behavior creates to the desire to conclude that these behaviors are acceptable, which initiates a search for reasons (which could also be found further back in time). In Chapter 6 we found that prior good deeds can serve as such a reason to give in to temptation.

As outlined in the introduction, definitions of self-licensing indicate that a prior good deed can *cause* undesirable behavior. We believe that temptation-based licensing is fundamentally different from this typical definition of good deed licensing. Whereas

good deed licensing implies that prior good behavior can create a *credit* to build a 'savings' account that one can spend on undesirable behavior (Miller & Effron, 2010), temptation-based licensing suggests that individuals who are tempted to display undesirable behavior will retrospectively search for a license. Thus, participants who experience temptation-based licensing do not have pre-earned credits.

We believe that both ways of self-licensing can be reflected in the research paradigms that are typically used to investigate the effect, where initial good behavior is manipulated and the effect of this manipulation on less desirable behavior is measured (e.g., Effron et al., 2009; Mazar & Zhong, 2010; Monin & Miller, 2001). The manipulations of good behaviors might have led to less desirable behaviors, as good deed licensing would predict. However, it could also be the case that participants felt tempted to engage in the "bad" behavior, and that this temptation made them search for a compelling reason, which was found in the salient prior good behavior. Either way, the phenomenon of self-licensing seems broader than the idea that a prior good deed leads to the undesirable behavior. Therefore, studies on self-licensing should also consider the possibility that the temptingness of the undesirable behavior can trigger the licensing process.

The possibility that self-licensing is initiated by the temptation to engage in certain behaviors is especially reflected in research on moral credentialing that suggests that people strategically attempt to earn moral credentials when they anticipate performing morally questionnable behaviors. For instance, participants who anticipated that their future behavior could be regarded as prejudiced exaggerated how favorably they perceived a black person in a previous job hiring task (Merritt et al., 2012). In a similar vein, Effron (2014) proposed that participants who were concerned that their future behavior could be regarded as prejudiced or unethical overestimated to what extent their previous non-racist choices or ethical behaviors proved their morality to other persons. Furthermore, in their research on counterfactual licensing, Effron, Monin, and Miller (2012; 2013) argued that in order to justify future undesirable behaviors (i.e., acting prejudiced and eating unhealthy foods), people tend to exaggerate negative counterfactuals of their foregone behavior. For instance, individuals who were

tempted to eat cookies rated previously declined foods as unhealthier compared to control participants who were not tempted. Thus, in these examples there was no direct good deed leading to a bad deed. Rather, the temptation to display undesirable behaviors made prior deeds seem as better reasons to give in to temptation. Further research should determine whether the licensed behaviors in these studies are indeed perceived as tempting as we would predict they likely are.

The findings of the current chapter have important implications for manipulations and measures used in self-licensing studies. Since the results of both studies are consistent with the idea that there are (at least) two different ways in which licensing can be triggered, researchers should carefully consider how they induce self-licensing. For instance, studies searching for moderators of the licensing effect predominantly focused on aspects of the prior good deed, but it can be very useful to focus on the temptingness of the undesirable behavior as well, given that licensing can also be elicited through the latter. Ultimately, emphasizing the role of the temptingness of the undesirable behavior in the emergence of the licensing effect can possibly explain why it is sometimes hard to replicate the 'traditional' good deed self-licensing effect (cf. Chapter 4): In some cases, the undesirable behavior that was measured in the paradigms may have not been that tempting for participants.

Implications

The distinction between the two ways of licensing gives important insights into the underlying processes of the self-licensing effect. Self-licensing theories and studies typically focus on aspects of the prior good deed, while for many instances of licensing people actually focus first and foremost on the desired bad deed. Thus, not only aspects of the prior good deed, but also tempting aspects of the undesirable behaviors are an important part of self-licensing theory.

We started this chapter with a short overview of the wide range of behaviors for which self-licensing can have adverse consequences. Our finding that a substantial part of self-licensing is triggered by the temptingness of the undesirable behavior can assist professionals in different domains (e.g., healthcare, ethics, etc.) in preventing such

adverse effects. Based on our results, we argue that at times when people are tempted to engage in these undesirable behaviors, justifying giving in through using prior good behaviors should become more difficult. For instance, health practitioners can use descriptive norms (Cialdini, Reno, & Kallgren, 1990) to inform people that healthy behaviors, such as working out and eating vegetables are normal rather than good (e.g., "75% of the people works out regularly"). This should be done at critical moments when people are tempted to indulge. In this way, the good behavior 'loses' its licensing capacities, which makes it harder to justify the undesirable behavior. Another way of making justifying through previous good deeds more difficult is to implement a system where people have to justify their choices to others. For instance, people who intend to save money can engage in a "commitment system" where they have to explain all their expenditures to other people.

Conclusion

In this chapter, we studied daily occurrences of self-licensing. This research has identified that there are at least two ways in which licensing can come about: *good deed self-licensing*, where licensing is triggered by aspects of the prior good deed, and *temptation-based self-licensing,* where licensing is triggered by the desired bad deed. We found that both ways have different antecedents and affective responses, highlighting the importance for researchers in the field to distinguish between these two ways of licensing. We hope that our findings stimulate further research on self-licensing and ultimately help in predicting when people display unfavorable or undesirable behavior.

Appendix 5.1. Categories of self-licensing

Categories of recalled self-licensing behaviors in Study 5.1

Category	Good behavior fell in this category	Licensed behavior fell in this category
Studying / working (vs relaxing / being lazy)	180 (64.5%)	122 (44.4%)
Working out / eating healthy (vs not working out / eating unhealthy)	54 (19.6%)	70 (25.5%)
Saving money (vs spending money)	6 (2.2%)	35 (12.7%)
Refrain from drinking alcohol /smoking (vs drinking and smoking)	4 (1.5%)	12 (4.4%)
Helping a friend	4 (1.5%)	0 (0.0%)
Other	14 (5.1%)	23 (8.4%)

Categories of recalled self-licensing behaviors in Study 5.2

	All participants *N* = 378		**Good Deed Licensing condition** *n* = 200		**Temptation-Based Licensing condition** *n* = 178	
	Good Behavior	Licensed behavior	Good behavior	Licensed behavior	Good behavior	Licensed behavior
Studying / working (vs relaxing / being lazy)	170 (45.0%)	204 (54.0%)	90 (45.0%)	108 (54.0%)	80 (44.9%)	96 (53.9%)
Working out / eating healthy (vs not working out / eating unhealthy)	82 (21.7%)	33 (8.7%)	43 (21.5%)	20 (10.0%)	39 (21.9%)	13 (7.3%)
Saving money (vs spending money)	24 (6.3%)	48 (12.6%)	9 (4.5%)	18 (9.0%)	15 (8.4%)	10 (16.9%)
Refrain from drinking alcohol / smoking (vs drinking and smoking)	12 (3.2%)	52 (13.8%)	2 (1.0%)	2 (1.0%)	10 (5.6%)	23 (12.9%)
Helping a friend (vs acting negatively towards others)	59 (15.6%)	3 (0.8%)	43 (21.5%)	29 (14.5%)	16 (9.0%)	1 (0.6%)
Other	31 (8.2%)	38 (10.1%)	12 (6.5%)	23 (11.5%)	18 (10.1%)	15 (8.4%)

CHAPTER 6

Temptation-Based Reasoning

Building on theories of motivated reasoning and reason-based choice, we propose and test a temptation-based reasoning model where people interpret reasons for indulgence in a different light depending on how tempting behavior is. We expected that reasons for indulgence are seen as more acceptable when behavior is more tempting. In two studies we indeed find that the more tempting behavior is, the more people evaluate given reasons as acceptable reasons to indulge (Study 6.1 and Study 6.2). Furthermore, Study 6.3 and 6.4 revealed that both recalled prior good behaviors and recalled prior frustrations are interpreted as good reasons to indulge when confronted with tempting behavior. This process of temptation-based reasoning can shed new light on existing theories on how people deal with temptations, notably those on self-licensing, comfort buying, and comfort eating.

This chapter is based on: Blanken, I., Van de Ven, N., & Zeelenberg, M. (2015c). Temptation-based reasoning. *Manusript under review.*

A temptation is, by definition, a strong urge to do or have something that also has negative consequences. Examples of temptations are: unhealthy, high caloric foods when one is dieting, attractive, interested people when one is in a committed monogamous relationship, checking social media during working hours, and spending money on luxurious items while trying to save money for retirement. Temptations thus represent a conflict between wanting something while preferring to resist it at the same time. Put differently, temptations represent a self-control problem. Much is learned over the last decades about how self-control may operate (Baumeister, Heatherton, & Tice, 1994; Baumeister, Vohs, & Tice, 2007; Fishbach, 2009; Muraven & Baumeister, 2000), but we think that an important aspect of temptations and how people deal with them is understudied. In the present chapter, we investigate what happens when people are confronted with temptations, as we believe that the temptation itself may already influence the thoughts and feelings of people, and these in turn influence the likelihood of acting on the temptation. We propose that the presence of a temptation causes people to look for reasons that allow them to act on it, that make giving in acceptable. In other words, temptations elicit a search for justifications. We believe that this also implies that the more tempting behavior is, the more people will regard any reason for giving in to temptation as an acceptable reason ("It is like putting butter in front of a cat"). People are thus expected to interpret reasons in a different light when confronted with a temptation.

This reasoning builds on various different lines of research. First, Kunda's (1990) theory on motivated reasoning posits that when individuals have a preferred conclusion, they are likely to construct justifications that are supportive of that particular conclusion. If we translate the ideas on motivated reasoning to temptations, this would imply that a temptation itself creates the desire to conclude that indulgence is acceptable, which in turn influences the reasoning process. People thus search for acceptable reasons to justify the desired indulgence. In a similar vein, research on moral reasoning shows that people first arrive at certain moral judgments, and subsequently try to construct reasons that support their judgments (Haidt, 2001).

Our reasoning also builds on reason-based choice research, which shows that many real life decisions are not necessarily based on a rational calculation of costs and benefits of taking an action, but on finding reasons or justifications that support one course of action above the other (Bastardi & Shafir, 2000; Shafir, Simonson, & Tversky, 1993). Thus, when people favor product A over product B, they will actively look for reasons that will justify this preference. For instance, physicians sometimes request additional tests that do not influence their treatment decisions, but provide them with extra reasons to justify the treatments they previously prescribed (Manktelow, 2012). In other words, actions for which reasons exist are more acceptable, and when one thus wants to act a certain way, finding reasons for doing so would help.

Our work further builds directly on the seminal work by Langer, Blank, and Chanowitz, (1978). When these researchers asked people using a copying machine if they could go first, the requests were typically denied. However, when they provided a reason for their request ('because I'm in a rush' or 'because I have to make copies') people complied and let them use the machine first. It thus seems that sometimes people simply need a reason or justification in order to do something.

Note that all these lines of research show that people like to find reasons for what they want to do. We extend these different lines of research to the field of temptations. We believe that when people are tempted to display certain behaviors, they search for justifications and any reason may serve as one. Importantly, we do not only believe that people who want to give in to temptation are more likely to find reasons (as theories on motivated reasoning and reason-based choice would predict), but, and this is where our contribution lies, we also expect that people regard a *similar* reason as more compelling when the temptation is stronger. Thus, we hypothesize that people would find a certain reason more acceptable when the temptation is stronger than when it is less strong. If externally provided or created reasons are actually regarded as better reasons to indulge, this could have important practical implications. For instance, providing in-store advertisements with a possible justification could be more effective for more tempting products.

The idea that the 'temptingness' of the behavior plays an important role in determining the acceptability of the reasons for indulgence finds initial support in recent research by De Witt-Huberts, Evers, and De Ridder (2014b). In their research, 58 female undergraduate students were asked to rate the temptingness of a chocolate bar. Next, these students saw a list of 30 reasons that could be provided for eating the chocolate, and they ticked the ones that applied to them. The more tempting they rated the chocolate, the more reasons the students ticked. In another study, 36 female undergraduate students were asked to generate and write down the reasons they would have to eat a similar chocolate bar. Again, the more tempting the chocolate bar was rated, the more justifications the students wrote down. Together, these results show a correlation between the self-rated temptingness of behavior and the number of justifications that people indicate or generate to justify giving in to this temptation. The causal direction of the effect is not clear yet, as the chocolate bar could also become more tempting when students are asked to provide more reasons for eating it (e.g., after working out, high calorie food might be more tempting as one needs to replenish one's energy level). To test our proposed model that people interpret reasons for indulgence in a different light when confronted with a temptation, it is crucial to experimentally manipulate the temptingness of the temptation and subsequently measure how this influences participants' reasoning processes.

If our prediction is correct that tempting situations make reasons to give in to these temptations seem more acceptable, this process can shed new light on existing theories. For example, an important theory about when people give into temptations is self-licensing (Miller & Effron, 2010). The traditional view of self-licensing is that a prior good deed allows someone to engage in subsequent behavior that one normally would refrain from. Various different theories on how the licensing process precisely operates exist (see Chapter 2 and Effron and Conway (2015) for reviews), but they all tend to focus on aspects of the prior good behavior that creates the license. For example, Khan and Dhar (2006) found that participants who first imagined performing voluntary work were subsequently more likely to prefer luxurious over practical sunglasses compared to participants who did not imagine performing voluntary work. They argued that (imagined) prior good behavior activates a positive self-concept,

which subsequently licenses the choice of a self-indulgent option (also see Dunning, 2007). Theorizing on self-licensing thus tends to focus on how the prior good deed affects subsequent behavior. However, if people regard reasons for indulgence as more compelling when behavior is more tempting, licensing research should in our view also focus on the possibility that it is mainly the temptation to engage in undesirable behavior that makes someone look for a compelling reason, and in typical licensing studies that might be the good behavior that has just been made salient by the set-up of the experiment. Thus, it could also be the case that participants in the Khan and Dhar study felt tempted to choose a luxurious over a practical pair of sunglasses, and that this temptation made them search for a compelling reason, which could be found in the saliency of the intention to perform a good deed.

Another domain in which the temptation-based reasoning model that we put forward here could be of importance is in comfort buying. Comfort buying or retail therapy refers to buying products as a way to alleviate negative moods (e.g., Garg & Lerner, 2013; Kacen, 1998; Rick, Pereira, & Burson, 2014). Atalay and Meloy (2011) found that negative moods lead to a greater consumption of unplanned self-treats. If our prediction is correct that any reason for indulgence is regarded as more compelling when behavior is more tempting, research on comfort buying could also focus on the possibility that perhaps the temptation to buy a self-treat makes someone look for a compelling reason, which could be found in the saliency of the current negative mood (e.g., 'These shoes are fabulous, but way beyond my budget. But I feel sad about failing my exam and think I deserve a break now and then').

In a similar vein, research on comfort eating – that is, eating to relieve negative emotions (Ganley, 1989) – could investigate the possibility that it is mainly the temptation to eat a mouthwatering snack that makes someone look for a compelling reason. For example, the finding that overweight individuals are more likely to overeat when feeling emotionally distressed than healthy-weight individuals (Arnow, Kenardy, & Agras, 1992; Baucom & Aiken, 1981; Chua, Touyz, & Hill, 2004; Lowe & Fisher, 1983; McKenna, 1972; Plutchik, 1976; Ruderman, 1983; Slochower, Kaplan, & Mann, 1981) may also occur if the food is more tempting for overweight individuals,

and, as a consequence, feeling emotionally distressed might be regarded as a more compelling reason for indulgence by those who find that type of food more tempting than others.

We would like to emphasize that we do not argue that our proposed process (a temptation triggers the search for a reason) is the only explanation for the above three examples (self-licensing, comfort buying, and comfort eating). We merely think that is a potential alternative explanation, in addition to the process put forward by the original authors (prior good deeds make questionable behavior more likely, distress causes comfort buying or comfort eating).

The studies in the current chapter

Taken together, the present chapter aims to shed light on when people permit themselves to give in to temptations. We believe that the presence of a temptation can initiate a motivated reasoning process, where reasons become more attractive for indulgence when the temptation is stronger. This adds to existing work because, 1) it shows that people do not only search harder to find reasons to reach a desired conclusion, but also think reasons are more acceptable so as to reach a desired conclusion, and because, 2) we manipulate the temptingness of possible indulgences, which allows us to establish the causal link as we propose it here. We first test this in two studies. Specifically, Study 6.1 tests whether a manipulation that influences the temptingness of checking social media on a personal smartphone during working hours affects how acceptable people find different reasons to check their phones. We predicted that when the smartphone was made more tempting, people would find any reason more acceptable to check the phone. Similarly, Study 6.2 tests whether the temptingness of a hamburger affects how acceptable people find different reasons to consume this burger.

In addition, we think that the process of temptation-based reasoning can account for established effects, such as self-licensing, comfort buying, and comfort eating. In Study 6.3 we test whether in a more tempting situation (the tempting burger again) people find a prior good deed a better reason to indulge, compared to in a less tempting

situation, testing whether self-licensing is influenced by the temptingness of the questionable behavior. In Study 6.4 we test whether people regard a prior frustration as a better reason to indulge in a more tempting situation, compared to in a less tempting situation, testing whether theories on indulgence to cope with negative emotions are influenced by the temptingness of the indulgence.

Study 6.1. The temptation to check social media during working hours

In Study 6.1, we investigated whether people find different reasons for checking social media on their personal smartphones during work more acceptable when there is a high temptation to check their smartphones. We manipulated temptation through showing pictures of a tempting smartphone screen with alerts and messages versus a less tempting smartphone screen without any alerts (Figure 6.1). We expected that participants who saw the picture of a tempting smartphone screen found various reasons to check their personal phones during work more acceptable than participants who saw the picture of a less tempting smartphone screen. In addition, we expected that the more tempted participants were by the pictures of the phone screen, the more acceptable they found different reasons to check the phone.

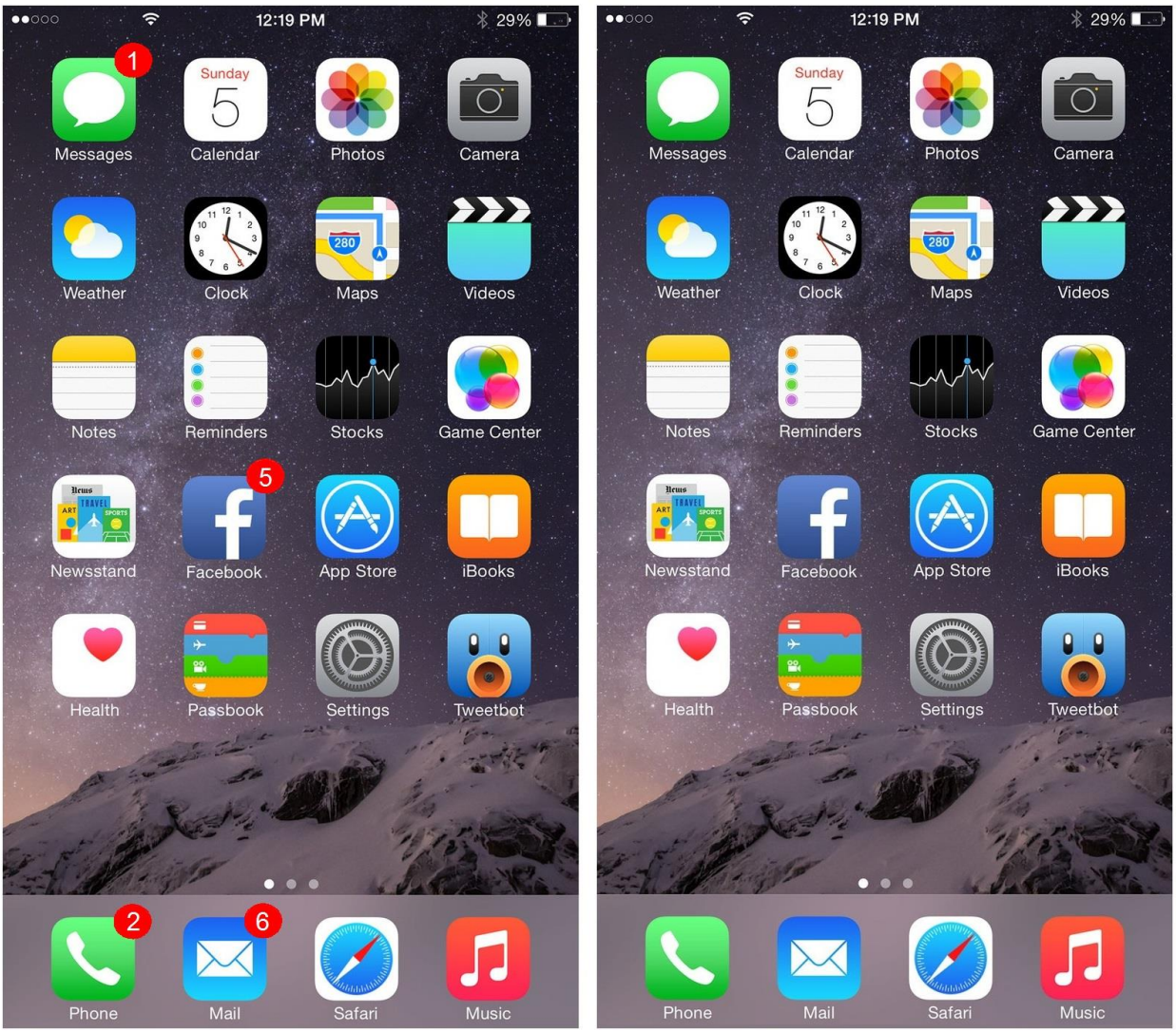


Figure 6.1. The smartphone screens shown in the Tempting Phone condition (left) and the Less Tempting Phone condition (right).

Method

We first conducted a pretest including 109 participants (65 males and 44 females, M_{age} = 32.11, *SD* = 10.54) to examine whether checking both phones during work time was regarded as equally undesirable. In this pretest, participants imagined working for a company that restricts the use of personal phones and saw the picture of the tempting smartphone (n = 55) versus the less tempting smartphone (n = 54). Subsequently, they were asked whether they thought that checking the phone during working hours would break their company policy (1 = *not at all*, 7 = *very much*). The results showed that checking the tempting smartphone was regarded as equally undesirable (M = 6.25,

SD = 1.27) as checking the less tempting smartphone was (*M* = 6.22, *SD* = 1.61, *t*(107) = 0.12, *p* = .907), indicating that for both phones, people realize that checking it during work hours is a negative action.

Four hundred and ninety-two U.S. based participants completed our study on Amazon Mechanical Turk (MTurk). We excluded participants who indicated that they did not own a smartphone (*n* = 34)[24], leaving 458 participants (272 males and 186 females, M_{age} = 31.05, *SD* = 9.14) in our sample. Participants were randomly assigned to the Tempting Phone condition (*n* = 221) or to the Less Tempting Phone condition (*n* = 237). Participants were exposed to one of the phone screen images (Figure 6.1) and read the following instructions:

> *Your work company has a policy that restricts the use of personal phones; in principle you are not allowed to check or use your personal phone during working hours. This is a screenshot of your personal smartphone that you use for private purposes only. You do not receive any work-related messages on this phone.*

Next, participants were asked to indicate how tempted they would be to check their personal phone if their screen looked like the presented image (1 = *not at all*, 7 = *very much*). Subsequently, participants were asked to read four potential reasons for checking this personal phone during work, and they indicated for each whether they found this reason an "acceptable" reason (1 = *not at all*, 7 = *very much*). The four reasons were: "You feel very frustrated about something that happened today", "The past few days, you have worked very hard; you are ahead of your working schedule", "You performed a good deed today, such as helping someone in need", and "It is a Tuesday afternoon".

Results

Participants in the Tempting Phone condition (*M* = 5.35, *SD* = 1.79) rated the phone screenshot as more tempting than participants in the Less Tempting Phone condition

[24] When we included participants who did not own a smartphone in the analyses, we found similar results.

(M = 4.08, SD = 2.05), $t(456) = 7.03$, $p < .001$, $d = 0.66$. Means, standard deviations and statistics of the acceptability of reasons are displayed in Table 6.1. A MANOVA including all reasons demonstrated that participants in the Tempting Phone condition regarded the reasons as marginally more acceptable than participants in the Less Tempting Phone condition, Wilks λ (4, 453) = .918, p = .073, η_{p^2} = .019. When analyzing the univariate results for each reason, we found that participants in the Tempting Phone condition regarded the frustration reason and the Tuesday afternoon reason as more acceptable reasons for checking their personal phones during work compared to participants in the Less Tempting Phone condition. There were no differences between conditions for the other two reasons. If we aggregate the acceptability of all four reasons into a general "reason acceptability"-score (α = .85), we see a significant effect of condition as well. For all reasons (including the reason acceptability at the aggregate level), the rated temptingness of the phones strongly predicted reason acceptability (all β's $\geq$ .27, t's (456) $\geq$ 6.03, p's < .001).

Table 6.1

Effects of phone condition (Tempting Phone versus Less Tempting Phone) on the acceptability of the specific reasons to check the phone during work in Study 6.1

	Phone				
	Tempting *n* = 221	**Less Tempting** *n* = 237			
Reason	*M (SD)*	*M (SD)*	*F*(1, 456)	*p*	η_{p^2}
You feel very frustrated about something that happened today.	3.76 (1.78)	3.41 (1.65)	4.93	.027	.011
The past few days, you have worked very hard; you are ahead of your working schedule.	4.78 (1.74)	4.65 (1.84)	0.67	.413	.001
You performed a good deed earlier today, such as helping someone in need.	3.51 (1.74)	3.27 (1.68)	2.19	.140	.005
It is a Tuesday afternoon.	2.99 (1.85)	2.59 (1.66)	6.07	.014	.013
Combined (α = .85)	3.67 (1.47)	3.48 (1.45)	4.41	.036	.010

Note. Acceptability of the reasons was rated from 1 *(not at all)* to 7 *(very much).*

Discussion

The findings of Study 6.1 show that the more tempted people are to check their personal phone, the more acceptable they find different reasons to check their phone. These findings are supportive of our hypothesis that people regard similar reasons as more compelling when the temptation is stronger. A possible reason why we did not find that all reasons were seen as more acceptable when the situation was more tempting, was that the temptingness manipulation only had a moderate effect size on how tempting it was to check the phone (d = 0.66). If we expect that the difference in perceived temptingness created by our manipulation would predict how acceptable

reasons are, the effect size of the temptingness manipulation on acceptability of reasons would, even with a perfect correlation of perceived temptingness with the acceptability of reasons, be constrained to that maximum effect size of $d = 0.66$. Given that the pattern of results we found in Study 6.1 is clear, but the effect not that strong (nor strongly significant), we wanted to replicate the initial findings. To strengthen our findings we attempted to create a stronger manipulation of perceived temptingness.

Study 6.2. The temptation to consume a tasty hamburger

The second study had two objectives. First, we aimed to investigate whether the findings from Study 6.1 applied more broadly and tested it within another domain of temptations (unhealthy snacking instead of questionable work behavior). The second objective was to use a stronger manipulation of temptingness. We investigated whether the confrontation with a hedonic food temptation leads people to find various reasons for indulgence more acceptable. We manipulated temptation through showing pictures of a tempting burger versus a less tempting burger (Figure 6.2). We expected that participants who saw the picture of the tempting burger found various reasons to consume this burger more acceptable than participants who saw the picture of the less tempting burger. In addition, we expected that the more tempted participants were by the pictures of the burger, the more acceptable they found different reasons to consume the burger.

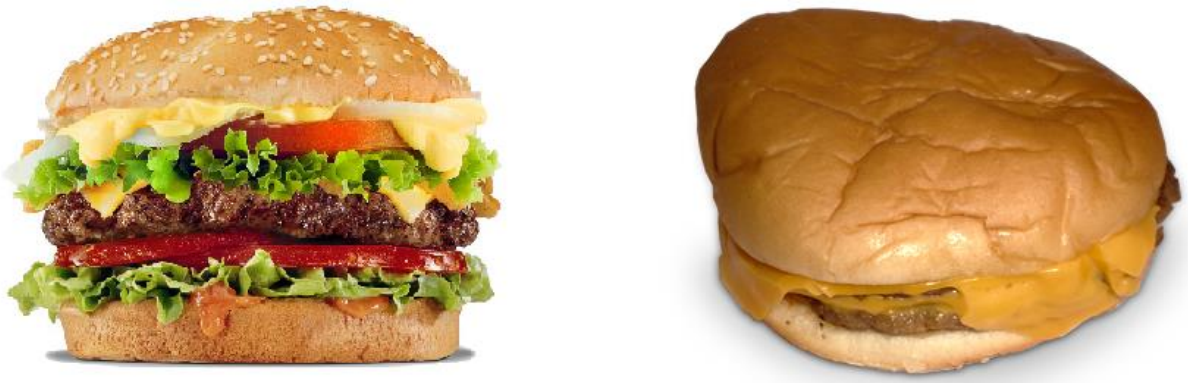


Figure 6.2. The burgers shown in the Tempting Burger condition (left) and the Less Tempting Burger condition (right)[25].

Method

We first conducted a pretest including 95 participants (56 males and 38 females (1 missing), M_{age} = 32.06, *SD* = 10.50) to examine whether eating both burgers was regarded equally undesirable. In this pretest, participants saw the picture of the tempting burger (*n* = 49) versus the less tempting burger (*n* = 46). Subsequently, they were asked how unhealthy they found this hamburger compared to other foods (1 = *not unhealthy at all*, 7 = *very unhealthy*). The results showed that eating the tempting burger (*M* = 5.33, *SD* = 1.16) and the less tempting burger (*M* = 5.70, *SD* = 1.19) was regarded equally undesirable, $t(93) = 1.53$, $p = .130$, indicating that in both conditions, people realize that eating the unhealthy burger is a negative action.

Five hundred and one U.S. based participants completed our study on MTurk. We excluded participants who indicated that they were vegetarians (*n* = 19)[26], leaving 482 participants (310 males and 172 females, *M*age = 31.70, *SD* = 10.91) in our sample.

[25] These images were derived from http://i.imgur.com/4hIyvq8.jpg?1 and http://i.imgur.com/eW8Z5rJ.jpg

[26] When we included participants who indicated that they were vegetarians, we found similar results.

Participants were randomly assigned to the Tempting Burger condition (n = 243) or to the Less Tempting Burger condition (n = 239) and were exposed to one of the different burgers presented in Figure 6.2. Participants were asked to take a close look at the burger and to indicate how tempting they found this burger (1 = *not at all*, 7 = *very much*). Participants then read that this burger was intended as an indulgence and that the producers of the burger wanted to know for what reasons individuals would allow themselves this particular burger (adopted from De Witt-Huberts et al., 2014b). Next, participants were asked to read four potential reasons for eating the burger, and they indicated for each whether they found this reason a "good reason" for them to eat it (1 = *not at all*, 7 = *very much*). The four reasons were: "Imagine that you had an intense workout at the gym today", "Imagine that it is a Friday afternoon", "Imagine that you have worked two hours on top of your normal working hours", and "Imagine that you feel very frustrated about a conflict with your coworker that happened earlier today".

Results

Participants in the Tempting Burger condition (M = 5.34, SD = 1.61) rated the burger as more tempting than participants in the Less Tempting Burger condition (M = 3.14, SD =1.84), $t(480) = 13.97$, $p < .001$, $d = 1.27$. The manipulation was intended to be stronger than the one in Study 6.1, which was successful (the manipulation of Study 6.1 had an effect size of d = 0.66 on the temptingness measure).

Means, standard deviations and statistics on the acceptability of reasons per condition are displayed in Table 6.2. A MANOVA including all reasons demonstrated that participants in the Tempting Burger condition found the reasons more acceptable than participants in the Less Tempting Burger condition, Wilks λ (4, 477) = .890, $p < .001$, η_{p^2} = .110. When analyzing the univariate results for each reason, we found that participants in the Tempting Burger condition regarded each reason as a better reason for consuming the burger compared to participants in the Less Tempting Burger condition. If we aggregate the acceptability of all four reasons into a general "reason acceptability"-score (α = .84), we see a significant effect of condition as well. For all reasons (including the reason acceptability at the aggregate level), the rated

temptingness of the burgers strongly predicted reason acceptability (all β's ≥ .43, *t*'s (480) ≥ 10.41, *p*'s < .001).

Table 6.2

Effects of burger condition (Tempting Burger versus Less Tempting Burger) on the acceptability of the specific reasons to eat the burger in Study 6.2

	Burger				
	Tempting *n* = 221	**Less Tempting** *n* = 237			
Reason	*M (SD)*	*M (SD)*	*F* (1, 480)	*p*	η_{p^2}
You had an intense workout at the gym today	3.98 (1.97)	3.10 (1.89)	24.80	<.001	.049
It is a Friday afternoon	4.42 (1.88)	3.23 (1.86)	49.24	<.001	.093
You worked two hours on top of your normal working hours	4.77 (1.78)	2.62 (1.94)	45.26	<.001	.086
You feel very frustrated about a conflict with your coworker that happened earlier today	3.61 (1.94)	2.67 (1.73)	31.47	<.001	.062
Combined (α = .84)	4.20 (1.43)	3.16 (1.57)	57.19	<.001	.106

Note. Acceptability of the reasons was rated from 1 *(not at all)* to 7 *(very much).*

Discussion

Study 6.2 replicated the findings of Study 6.1 in the domain of unhealthy snacking: Participants in the Tempting Burger condition found all reasons more acceptable for indulgence compared to participants in the Less Tempting Burger condition. In addition, the results reveal that the more tempting people find a burger, the more acceptable they find different reasons to eat this burger.

Together, Study 6.1 and 6.2 show a clear relation between how tempting something is, and whether diverse reasons for displaying this behavior are acceptable: When a temptation is more tempting, people find the same reason to give in to that temptation more acceptable than when the temptation is less tempting. Note that in both cases

participants would realize that their behavior is undesirable: the pretest shows that people think both burgers are equally unhealthy.

Interestingly, participants in the conditions with more tempting behavior sometimes even judged ostensibly unrelated reasons as more acceptable reasons than participants in the conditions with less tempting behavior. For example, 'it is a Friday afternoon' was regarded as a more acceptable reason to consume a burger when this burger looked very attractive. This illustrates that even reasons that convey no relevant information are more acceptable in front of a temptation. This is consistent with the findings of Langer et al. (1978) that the content of a reason can be irrelevant for acting upon that reason.

So far, we only investigated the acceptability of reasons we provided to participants. In addition, participants read about the different reasons directly after being exposed to the temptation. In Study 6.3 and 6.4, we investigate whether self-generated reasons prior to exposure to a temptation are interpreted differently depending on the temptingness of the temptation. We think this process has important implications for theories on self-licensing (as we will explain in Study 6.3) and comfort buying and eating (as we will explain in Study 6.4).

Study 6.3. The interpretation of previous good behaviors vis-à-vis a temptation

We next investigated whether people find their own previous good deeds more acceptable reasons to indulge when they are exposed to a tempting burger versus a less tempting burger. So far, we have found the pattern of results that suggest that any reason is seen as a better reason to indulge when confronted with more tempting behavior. We therefore also expect that a prior good deed would be seen as a better and more acceptable reason to indulge with a stronger temptation. This resembles the process of self-licensing, the empirical finding that people are more likely to engage in questionable behavior after having performed a good deed (Miller & Effron, 2010). The process of self-licensing is typically perceived as one in which the initial good act later influences people's behavior: "virtuous acts can license subsequent asocial and

unethical behaviors" (Mazar & Zhong, 2010, p. 495). However, given our results of Study 6.1 and 6.2, we think it also likely that the process works in the reverse order as well: When confronted with a temptation, people look for an acceptable reason to give in to that temptation, and a prior good deed is more easily accepted as a good reason when the temptation is stronger.

To summarize, we expected that participants who saw the picture of a tempting burger would find their previously recalled good deeds more acceptable reasons to eat this burger than participants who saw the picture of a less tempting burger. In addition, we expected that the more tempted participants were by the pictures of the burger, the more they would regard their previous good deed as an acceptable reason to eat the burger.

Method

Five hundred and one U.S. based participants completed our study on MTurk. We excluded participants who indicated that they were vegetarian (n = 25)[27], leaving 476 participants (311 males and 165 females, Mage = 30.96, SD = 10.29) in our sample. Participants were randomly assigned to the Tempting Burger condition (n = 234) or to the Less Tempting Burger condition (n = 242). Participants read the following instructions:

> *Think about a recent situation in which you performed a good deed. By a good deed, we mean a situation where you displayed good, moral, or virtuous behavior. Please describe this good deed.*

Next, participants saw a picture of either the tempting or the less tempting burger from Study 6.2. Participants were asked to take a close look at the burger and to indicate how tempting they found this burger (1 = *not at all*, 7 = *very much*). Participants then read that this burger was intended as an indulgence and that the producers of the burger wanted to know for what reasons individuals would allow themselves this

[27] When we included participants who indicated that they were vegetarian, we found similar results.

particular burger (adopted from De Witt-Huberts et al., 2014b). Next, the good deed that the participants previously recalled was displayed on the screen. Participants were asked whether they found their own good deed a good reason to eat the burger (1 = *not at all*, 7 = *very much*).

Results

Participants in the Tempting Burger condition rated the burger as more tempting (M = 5.54, SD = 1.62) than participants in the Less Tempting Burger condition (M = 2.97, SD =1.73), $t(474) = 16.15$, $p < .001$, $d = 1.53$. Participants in the Tempting Burger condition (M = 3.22, SD = 2.05) regarded their own prior good deed as a better reason for eating the burger compared to participants in the Less Tempting Burger condition (M = 2.28, SD = 1.69), $t(474) = 5.47$, $p < .001$, $d = 0.50$. Like we found before, the more the burger was perceived as tempting, the more someone thought their prior good deed was an acceptable reason to indulge ($\beta = .46$, $t(474) = 11.35$, $p < .001$).

Discussion

Together, these results suggest that even reasons that are generated prior to exposure to the temptation can serve as acceptable justifications to eat the burger, depending on how tempting people find this burger. A recalled prior good deed is thus seen as a better reason to indulge when a temptation is really tempting, than when it is less tempting.

This finding may closely resemble what happens when people display self-licensing effects, where initial good or moral behavior leads to behavior that is immoral, unethical, or otherwise problematic (Miller & Effron, 2010). Research on self-licensing predominantly focused on the role of previous good behavior in leading to unwanted behavior. For example, Fishbach and Dhar (2005) found that participants who first perceived that they made progress on their weight loss objectives were later more likely to choose a chocolate bar over an apple as a participation gift. Our findings imply that licensing research could also focus on the possibility that it is the temptation to indulge that makes someone look for a compelling reason to do so. Thus, it could also be the case that participants in the Fishbach and Dhar study felt tempted to choose a

chocolate bar over an apple, and that this temptation made them search for a compelling reason. In other words, a prior good performance is perhaps easier to accept as a good reason than a recalled neutral situation (their control condition).

As we explained in the introduction, we think that these results are not limited to recalls of prior good behavior. If 'any' reason serves as a license to give in to temptation, we can expect similar effects for recalls of prior *negative* behavior. For example, having a bad day can also serve as a good excuse to indulge when something is really tempting (i.e., comfort buying or comfort eating). Therefore, in Study 6.4, we investigate whether recalls of recent frustrations are interpreted differently depending on the temptingness of the temptation.

Study 6.4. The interpretation of previous frustrations vis-à-vis a temptation

In Study 6.4, we investigated whether people find their previous frustrations more acceptable reasons to indulge when they are exposed to a tempting burger versus less tempting burger. We expected that participants who saw the picture of a tempting burger would find their previously recalled frustrations more acceptable reasons to eat this burger than participants who saw the picture of a less tempting burger. In addition, we expected that the more tempted participants were by the pictures of the burger, the more they would regard their previous frustration as an acceptable reason to eat the burger.

Method

Four hundred and eighty seven U.S. based participants completed our study on MTurk. We excluded participants who indicated that they were vegetarian (n = 24)[28], leaving 463 participants (279 males and 184 females, *M*age = 31.54, *SD* = 9.91) in our sample. Participants were randomly assigned to the Tempting Burger condition (n =

[28] When we included participants who indicated that they were vegetarians, we found similar results.

233) or to the Less Tempting Burger condition (n = 230). Participants read the following instructions:

> *Think about a recent situation in which you felt very frustrated. Please describe this situation.*

Next, participants saw a picture of either the tempting or the less tempting burger from Study 6.2. Participants were asked to take a close look at the burger and to indicate how tempting they found this burger (1 = *not at all*, 7 = *very much*). Participants then read that this burger was intended as an indulgence and that the producers of the burger wanted to know for what reasons individuals would allow themselves this particular burger adopted from De Witt-Huberts et al., 2014b). Next, the frustration that the participants previously recalled was displayed on the screen. Participants were asked whether they found their own frustration a good reason to eat the burger (1 = *not at all*, 7 = *very much*).

Results and discussion

Participants in the Tempting Burger condition (M = 5.26, SD = 1.72) rated the burger as more tempting than participants in the Less Tempting Burger condition (M = 3.10, SD = 1.83), $t(461) = 13.11$, $p < .001$, $d = 1.22$. Participants in the Tempting Burger condition (M = 3.11, SD = 2.31) regarded their own frustration as a better reasons for eating the burger compared to participants in the Less Tempting Burger condition (M = 2.31, SD = 1.83), $t(461) = 4.36$, $p < .001$, $d = 0.41$. The more the burger was perceived as tempting, the more someone thought their prior frustration was an acceptable reason to indulge ($\beta = .44$, $t(461) = 10.61$, $p < .001$).

These results show that unrelated frustrations can also serve as justifications to eat the burger, depending on how tempting people find this burger. Combined with the results of Study 6.3, these findings support our theorizing that 'any' reason can serve as a license to indulge when confronted with a temptation. We will come back to the general consequences of this finding for theories on comfort buying and eating in the General Discussion section.

General Discussion

We proposed a temptation-based reasoning model where the presence of a temptation can cause people to look for reasons to act on it. Our studies show that people find a large variety of reasons for indulgence more acceptable when exposed to a tempting situation compared to a less tempting situation (Study 6.1 and Study 6.2). In addition, our findings show that both recalled prior good deeds (Study 6.3) and frustrations (Study 6.4) are interpreted as better reasons to indulge when confronted with a temptation. The finding that temptations are important for the interpretation of different reasons for indulgence has important implications for theories on how people deal with temptations (notably self-licensing theory and theories on comfort buying and eating). We will first discuss the importance of our account of tempted reasoning, after which we discuss the implications of our findings for other theories.

Tempted reasoning

The present findings build on theories of motivated reasoning (Kunda, 1990) and reason-based choice (Bastardi & Shafir, 2000; Shafir et al., 1993), which posit that when people want to reach a certain conclusion they will construct supportive reasons for doing so. Our findings show that people who want to give in to temptation are not only more likely to find reasons for indulgence (see De Witt-Huberts et al., 2014b), but, importantly, they also regard *the same* reason as more compelling when the temptation is stronger. Consistent with the findings of Langer et al. (1978) that people are more willing to comply with a request in the presence of a reason - regardless of its content -, we found that participants in the conditions with more tempting behaviors judged ostensibly unrelated reasons as more acceptable for indulgence compared to participants in the conditions with less tempting behaviors. For example, 'it is a Tuesday afternoon' becomes a better reason for checking a personal phone at work when checking this phone is more tempting. Thus, people need a reason to give in to temptation, but they do not care about the contents of that reason.

Of course, not all reasons are created equally. With which we mean that, although we have found that temptation tends to make all reasons somewhat more acceptable, it

does not mean that all reasons will become acceptable enough to act on the indulgence. It may as well be the case that the reason "It is Tuesday afternoon" is seen as a better reason when facing a stronger temptation, but it might still not be seen as a good enough reason to actually indulge and give in to the temptation. Our case for temptation-based reasoning does not imply that any reason is good enough to indulge, but that temptations will increase the acceptability of reasons for the indulgence and thereby make the indulgence more likely. It will still be the case that some reasons are simply better than others.

Self-licensing

Study 6.3 shows that previous good deeds are seen as a better reason to indulge when the temptation is stronger. This finding is very interesting in the perspective of research on self-licensing, the phenomenon that after displaying good behavior, people feel licensed to display undesirable behaviors (Miller & Effron, 2010). The assumed process is that self-licensing consists of two consecutive behaviors or events, where initial good behavior leads to less desirable behavior. Miller and Effron for example argue that the prior good deed creates either a credit (something that can later be used to trade in to transgress) or serves as a credential (the good deed builds a positive reputation, after which a transgressions is seen as less bad because of one's good reputation). In both these cases it is the prior good deed that builds a credit or a credential, which later has effects. This reflects how self-licensing is typically investigated in the lab; initial good behaviors are manipulated and the effects on less desirable behaviors are measured. For instance, recalls of good or moral behavior (Conway & Peetz, 2012; Jordan, Mullen, & Murnighan, 2011; Sachdeva, Iliev, & Medin, 2009) were used to manipulate the effect. However, it has not explicitly been tested whether the process of self-licensing necessarily operates in the order of these two consecutive behaviors. Our results suggest that the opposite process also works: A temptation triggers a search for acceptable reasons to transgress, and a prior good deed is seen as an acceptable reason to do so.

Note that the effects of moral licensing (i.e., psychological licensing in the moral domain) turn out not to be as robust as previously assumed. For instance, we could

not replicate the original findings of Sachdeva et al. (2009) that writing about one's positive traits leads to lower donations to charity and decreased cooperative behavior in a commons dilemma (cf. Chapter 4). In addition, the meta-analysis from Chapter 2 revealed that the average effect size of moral licensing is small-to-medium (Cohens *d* = 0.31). Moreover, the *p*-curve analysis from Chapter 3 indicated that research on moral licensing lacks evidential value. Since the findings in the current chapter show that temptations play an important role in the interpretation of previous behaviors, the licensing effect may be stronger and more robust when the undesirable behavior is more tempting.

Our findings on tempted reasoning thus serve as an initial step toward an alternative account of licensing where the temptingness of the undesirable behavior initiates a search for a license to indulge. This implies that it may not always be the case that after displaying good or moral behavior, people engage in less desirable behaviors. However, if one needs a license to give in to the temptation of undesirable behavior, it is likely that one will find such a license. This license can consist of previous good behaviors, but, as Study 6.4 shows, prior negative behavior or feelings can also serve as a license.

We do not think or argue that all licensing effects occur via this justification process, but do believe that the current literature typically interprets licensing as a good deed leading to a bad deed. The reverse process (a temptation making a prior good deed seem as a better reason to indulge) may actually be just as likely to occur. As explained before, this is a slightly different view of licensing, but an important distinction to make as for typical licensing theories and studies the focus seems to lie on aspects of the prior good deed, while we think that for many examples of licensing the person actually focuses first and foremost on the desired bad deed. For instance, we think it is not necessarily the case that after working out or eating healthy, one feels that one has earned a license to eat unhealthy food. However, we do think it is very likely that the moments when people are tempted to eat unhealthy foods will elicit a search for justifications, which could be found in prior food restraints and prior workouts.

Comfort buying and comfort eating

The finding from Study 6.4 that previous frustrations are regarded as more compelling reasons for indulgence in front of a tempting burger is very interesting in the perspective of research on comfort buying and comfort eating. Similar to research on self-licensing, research in these domains mainly focuses on previous negative affective states rather than the temptingness of the (food) products. For instance, it has been found that negative moods can lead to the purchase of self-treats (Atalay & Meloy, 2011). In a similar vein, negative emotions such as sadness can trigger binge eating in obese individuals (Chua et al., 2004). Our findings imply that research could also focus on the possibility that it is the temptation to eat unhealthy foods or to buy self-treats that makes someone search for a compelling reason to indulge, and find prior negative feelings as a reason for doing so. Put it differently, participants who feel tempted to indulge may search for compelling reasons to justify indulgence, which can be found in their (previous) affective states.

If our findings partially explain comfort buying and eating, interventions that focus on restoring someone's self-esteem or improving someone's mood (e.g., Schmeidel & Vohs, 2009) might not be fully effective for reducing these behaviors when the process we document in the current chapter is at play as well. Improving someone's mood might just remove the prior negative feeling as a reason to indulge, but replace it with a positive mood that could serve as a reason to indulge as well. Interventions aimed at making temptations less strong might be more effective then. We do not wish to claim that interventions aimed at improving mood or self-esteem do not work, but do think the current chapter helps in identifying boundary conditions in which they are most likely to be effective.

Conclusion

The studies in the current chapter reveal a temptation-based reasoning process, in which the temptingness of the temptation strongly influences how people interpret different reasons for indulgence. The contents of these reasons do not seem that relevant: If something is very tempting, 'any' reason becomes seen as a better reason

to give in to temptation. These findings are not only important for our understanding of how people deal with temptations, but also have important implications for other theories. Notably, our findings suggest an alternative account of self-licensing where the temptingness of the undesirable behavior initiates a search for a license, instead of the previous good behavior making subsequent undesirable behaviors more likely. A similar account can be proposed for findings on comfort buying and eating, where the temptingness of the indulgence can initiate a search for a compelling reason for indulgence.

CHAPTER 7

Reasonable Reasons for Indulgence

Previous research identified the most common reasons people have for breaking their diet. In the current chapter we elaborate on this by shifting the focus from the quantitative aspects of the reasons for breaking a diet (which reasons are most common?) to the qualitative aspects (what are the consequences of breaking a diet for various reasons?). We expected that the types of reasons that people have for breaking their diet play a key role in determining whether they will be motivated to resume their diets. We found that reasonable reasons for unhealthy snacking have less negative consequences for goal motivation than unreasonable reasons for unhealthy snacking (Study 7.1). Furthermore, Study 7.2a and 7.2b revealed that reasons for unhealthy snacking that have been identified in the literature differ in how reasonable they are. Moreover, Study 7.3 shows that the two categories of reasons for unhealthy snacking that are most common (enjoying a special occasion versus opportunity-induced eating) have different motivational consequences. Together, our results show that some reasons for breaking one's diet may actually be good reasons that do not result in a loss of motivation.

This chapter is based on: Blanken, I., Van de Ven, N., Spälti, A. K., & Zeelenberg, M. (2015). Reasonable reasons for indulgence. *Manuscript under review*.

One of the problems dieters face is the frequent exposure to tempting high caloric foods. Such encounters will sometimes lead to dieting failures, when people give in to temptation. These failures have immediate implications for their dieting goals as one consumes more calories than one should have. But giving in to temptation can also have even more serious and longer term consequences. Goal violations can result in motivation loss, negative emotions, and a poorer goal performance (Soman & Cheema, 2004). Specifically, violating one's diet can produce a "What the Hell effect," in which people abandon a dieting goal after eating a "forbidden" snack (Cochran & Tesser, 1996; Polivy & Herman, 1985). To summarize, breaking one's diet does not only lead to a higher intake of calories, it can also make people lose their motivation to continue dieting. As yet, however, it is not clear whether all goal violations have similar detrimental effects. In the current chapter, we investigate whether and when people lose their motivation after breaking their diet.

We believe that the negative effects of breaking one's diet may vary. We think this is the case because people have different reasons for breaking their diet. Some reasons may actually be considered "good" reasons that we expect not to result in a loss of motivation. You may be on a diet, but when your best friend is getting married it is not so bad to eat part of the wedding cake, and you easily pick up your diet the next day.

When facing a conflict between an immediate temptation and a long-term goal, people often seek and construct reasons to make their decisions or to justify them after having made the choice (Shafir, Simonson, & Tversky, 1993; Kivetz & Simonson, 2002; Kivetz & Zheng, 2006). Recent research identified the specific common reasons that people provide for the consumption of unhealthy foods. Taylor, Webb and Sheeran (2014) asked 371 student participants to rate the prevalence of various justifications 'just before eating something unhealthy'[29]. They found six broad classes of reasons or

[29] Taylor et al. developed these justifications in a pilot study with two focus groups (*n* = 10 and 6). Participants in these focus groups were asked to recall instances from the previous month when they had made a deliberate decision to eat something unhealthy and then to report what they had thought to themselves just before doing

justifications (in sequence of how common the reasons were): (1) the easy availability of the food, (2) compensatory behaviors (e.g., making up for snacking by exercising later), (3) exceptions to the norm (e.g., once in a while is OK), (4) deservingness (where the unhealthy food serves as a reward), (5) curiosity (e.g., it looks gorgeous), and (6) irresistibly (i.e., the hedonic aspects of unhealthy food such as taste and smell). They also found a correlation between the number of justifications used and the number of unhealthy snacks consumed, in that the more reasons one had for indulging the more one indulged. This is consistent with the idea that justifications are recruited to rationalize goal violations. Put differently, when people find a reason to snack, they are more likely to do so.

In a similar vein, Verhoeven, Adriaanse, De Vet, Fennis and De Ridder (2015) asked 1544 members of the general public to fill out the Reasons to Snack Inventory[30]. They detected six categories in the reasons that people provide for engaging in unhealthy snacking (again in sequence of how common they were): (1) to enjoy a special occasion, (2) opportunity-induced eating, (3) to gain energy, (4) to reward oneself, (5) because of social pressure, and (6) to cope with negative emotions. Thus, both Taylor et al. (2014) and Verhoeven et al. find that the a-typicality of the situation (e.g., 'I have something special to celebrate') and the availability of food (e.g., 'the food looks good') are among the predominant reasons that people use for unhealthy snacking. Importantly, Verhoeven et al. argued that health interventions should specifically target the reasons that people most frequently use to justify unhealthy snacking, because these are the main triggers for this undesirable behavior. Although Verhoeven et al. convincingly showed what the most common reasons for unhealthy snacking are, we are not yet convinced that these are the best reasons to target in health interventions.

---

so. An additional group (n = 15) then rated how frequently they used each type of justification.

[30] The Reasons to Snack inventory was developed from reasons that were identified in previous studies assessing reasons for unhealthy snacking (total n = 525), 18 items from Jackson, Copper, Mintz, & Albino (2003) and five items that were added by the authors.

When designing health interventions, we propose that it is not only important to focus on the most common reasons for unhealthy snacking, but rather on those that have the largest negative effects on further goal motivation. Verhoeven et al. (2015) explored the categories of reasons that people use to justify incidental snacking, but did not investigate the motivational consequences of having violated one's diet for reasons from these different categories. Taylor et al. (2014) found that the use of reasons can undermine participants' dietary intentions. However, since they manipulated having justifications in general and did not distinguish between the different types of reasons, it was not possible to draw conclusions on the consequences of specific types of reasons.

The present chapter examines the motivational consequences of the different reasons that people provide for breaking their diet. We expect that the reasons that people have for breaking a diet differ in reasonability, that is, how good people perceive them to be to break their diet. Some reasons are simply better reasons than other reasons. Breaking your diet during the dinner of your best friend's wedding seems a good reason to us. Breaking your diet because a cake looks delicious does not seem like a good reason at all. It is further expected that bad reasons for breaking one's diet will have the most negative consequences for goal motivation.

We believe that breaking one's dieting goal for a bad reason can influence the belief in one's own self-control capacities (Muraven & Baumeister, 2000). Individuals who provide bad reasons for unhealthy snacking may experience that they lack self-control and therefore perceive their violation as a failure. Since prior goal failures can result in decreased motivation (e.g., Louro, Pieters, & Zeelenberg, 2007; Nelissen, De Vet & Zeelenberg, 2011), individuals who provide bad reasons for indulgence may feel that their goal is not achievable to them.

In addition, if individuals perceive that the reasons they have for breaking their diet are not reasonable (e.g., I ate the cake because it looked delicious), this may signal that their goal is not important enough to them. Since perceived goal importance is a strong predictor of goal performance (Hollenbeck & Williams, 1987), we expect that individuals who provide bad reasons for indulgence will experience lower goal

commitment in the long run. In contrast, having a good reason (e.g., I ate the cake because it was my friend's wedding) may inoculate one against the negative impact of goal violations. Therefore, we expect that if one has a good reason to violate one's diet, the violation would only have little or no impact, and people would find their dieting goal as important as before. Note that the two reasons from the examples (opportunity-induced eating and enjoying a special occasion) are among the most common reasons for breaking a diet, but we expected that only the bad reason would have longer term negative consequences. Thus, how good a reason for breaking one's diet is can affect both goal feasibility and goal commitment. Therefore, we predict that, compared to good reasons, bad reasons for indulgence would have more detrimental effects on one's diet.

So far, research on the reasons people have for snacking has predominantly focused on the short-term negative effects of these reasons on dieting. The research in the present chapter aims to shed some light on how the different reasons that people provide for breaking their diet affect further goal commitment. We will identify how reasonable people find different reasons for breaking their diet. In addition, we will examine whether good reasons for indulgence have less negative consequences for long-term goal motivation than bad reasons for indulgence. A better understanding of these motivational consequences can assist in developing health-related interventions that target the reasons that have the most detrimental effects on weight loss goals.

Study 7.1. Providing reasons for indulgence

In order to examine whether the reasonability of a reason for breaking a dieting goal (i.e., how good or bad a reason is) plays an important role in determining whether people regain motivation to continue with their diet, participants were asked to generate a good versus a bad reason for eating French fries while on a diet. We expected that participants who imagined breaking their diet for a good reason would subsequently display more goal motivation, higher estimated goal feasibility, and higher perceived goal importance, compared to participants with a bad reason. Note that the infringement on the diet is the same in both conditions, but only the reasons for indulgence changed.

Method

Two hundred and seventy Tilburg University undergraduate psychology students[31] participated in a 60-minute research session of unrelated experiments (71 males and 199 females, M_{age} = 20.12, *SD* = 2.42). Participants were randomly assigned to the Good Reason condition (*n* = 138) or to the Bad Reason condition (*n* = 132). Participants were asked to imagine the following situation:

> *You recently gained weight. Your clothes are too tight and you are not feeling well. Therefore, you start dieting. You decide to stop eating unhealthy snack foods. A couple of days after starting your diet, you walk across a fast-food restaurant during dinnertime. You smell the mouth-watering aroma of French fries. You are hungry and are craving for a portion of French fries.*

Participants in the Good Reason condition were then asked to provide a reason that made eating the French fries in this particular situation acceptable to them, whereas participants in the Bad Reason condition were asked to provide a reason that was not acceptable.

Next, all participants imagined eating the French fries for the reason that they provided, after which they responded to questions tapping goal motivation, goal feasibility, and goal importance (1 = *not at all*, 7 = *very much*). Goal motivation and goal feasibility were measured by four different items; 'After eating the fries, would you feel that your diet is ruined?', 'After eating the fries, would you feel that there is not much point in continuing your diet?', 'After eating the fries, how easy would it be to resume your diet?', and 'After eating the fries, how attainable would your dieting goal be for you?' and goal importance via 'After eating the fries, how important would your dieting goal be for you?' (1 = *not at all*, 7 = *very much*).

[31] Ninety-one of those students indicated that they were currently on a diet. The results did not differ between dieters versus non-dieters.

Results and discussion

Participants in the Good Reason condition often provided reasons related to special occasions ('because it is my birthday', 'because I want to celebrate passing my exam') previous healthy behaviors ('because I just went to the gym') or their current energy levels ('because I would otherwise faint'). Participants in the Bad Reason condition often provided reasons related to laziness ('because I am too lazy to make dinner myself'), hunger ('because I am really hungry'), or simply feeling like eating a snack ('I am really craving for French fries').

Participants in the Good Reason condition were more motivated to continue with their diet and found their dieting goal more feasible compared to participants in the Bad Reason condition. In addition, participants in the Good Reason condition indicated that their dieting goal would be more important to them than participants in the Bad Reason condition. Means, standard deviations, and statistics are displayed in Table 7.1.

Table 7.1

Means, standard deviations, and statistics of goal motivation, goal feasibility, and goal importance of the Good and Bad Reason conditions in Study 7.1

	Good Reason condition *n* = 138	**Bad Reason condition** *n* = 132			
Measure	*M (SD)*	*M (SD)*	*F*	*p*	η_{p^2}
After eating the fries, would you feel that your diet is ruined?	3.72 (1.68)	5.23 (1.40)	63.52	<.001	.19
After eating the fries, would you feel that there is not much point in continuing your diet?	2.74 (1.45)	4.05 (1.81)	43.40	<.001	.14
After eating the fries, how easy would it be to resume your diet?	4.92 (1.51)	3.82 (1.48)	36.74	<.001	.12
After eating the fries, how attainable would your dieting goal be for you?	5.22 (1.26)	4.31 (1.39)	32.26	<.001	.11
After eating the fries, how important would your dieting goal be for you?	5.59 (1.27)	5.84 (1.43)	19.69	<.001	.07

Note. All variables ranged from 1 *(not at all)* to 7 *(very much).*

These results provide initial support for our theorizing that the reasonability of a reason for breaking a diet plays a crucial role in further goal striving: People who provide good reasons for indulgence do not lose motivation to resume their diet, compared to people who provide bad reasons. At first sight, this may seem intuitive, but note that the implications of this finding for health interventions could be far-reaching. Verhoeven et al. (2015) indicated that health interventions should focus on the most common reasons, while we believe we should also focus on those with the worst consequences (the bad reasons). We continued our investigation by examining

whether the various reasons for unhealthy snacking identified in the literature (Taylor et al., 2014; Verhoeven et al.) may differ in reasonability.

Study 7.2a and 7.2b. The reasonability of reasons for indulgence

Method

The participants for Study 7.2a and 7.2b were recruited from Amazon Mechanical Turk (MTurk). One hundred and fifty-two U.S. based participants (85 males and 67 females, M_{age} = 34.38, *SD* = 11.30) completed the survey for Study 7.2a. Study 7.2b was completed by 149 participants (88 males and 61 females, M_{age} = 33.07, *SD* = 11.22).

In both studies, the participants were asked to evaluate the different reasons for unhealthy snacking using seven-point Likert scales (1 = *not at all acceptable*, 7 = *very acceptable*). For Study 7.2a, the 22 reasons for unhealthy snacking from the Reasons to Snack inventory by Verhoeven et al. (2015) were measured for acceptability by all participants. As in Verhoeven et al. the relevant items were averaged to create the six subscales for unhealthy snacking: (1) to enjoy a special occasion (α = .83), (2) opportunity-induced eating (α = .82), (3) to gain energy (*r* = .43), (4) to reward oneself (α =.84), (5) because of social pressure (α = .73), and (6) to cope with negative emotions (α = .90).

For Study 7.2b, the participants evaluated the 54 reasons from Taylor et al. (2014). The relevant items were averaged to create the six categories of reasons for unhealthy snacking: 1) the availability of the food (α = .90), (2) compensatory behaviors (α = .96), (3) exceptions to the norm (α = .90), (4) deservingness (α = .94), (5) curiosity (α = .92), and (6) irresistibly (α = .92). In both studies, the order of all items was randomized. Finally, participants indicated whether they were on a diet, if they were trying to lose weight, and if eating healthy was important to them[32].

[32] In Study 7.2a, thirty-six participants (23.7%) indicated that they were currently on a diet, 71 (46.7%) indicated that they currently had the goal of losing weight. On average, participants found it quite important to eat healthy (low-calorie) food, $M_{importance}$ = 4.62 (*SD* = 1.61). In Study 7.2b, forty-nine participants (32.9%) indicated that they were currently on a diet, 80 (53.7%) indicated that they currently had the

Results Study 7.2a

An ANOVA with the perceived acceptability of each of the categories of reasons from the Reasons to Snack inventory (Verhoeven et al., 2015) was conducted with the categories as a within-subjects factor. Maulchy's test indicated that the assumption of sphericity had been violated, $\chi^2(14) = 71.23$, $p < .001$, therefore the degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = 0.85$). The results showed that the six categories of reasons for snacking varied significantly in their perceived acceptability, $F(4.25, 641.02) = 88.04$, $p < .001$, $\omega^2 = 0.33$. Means, standard deviations, and the results of the LSD post hoc comparisons are displayed in Figure 7.1.

goal of losing weight. On average, participants found it quite important to eat healthy (low-calorie) food, $M_{importance} = 4.59$ ($SD = 1.74$). In both studies, these factors did not affect responses and were therefore not included in the analyses.

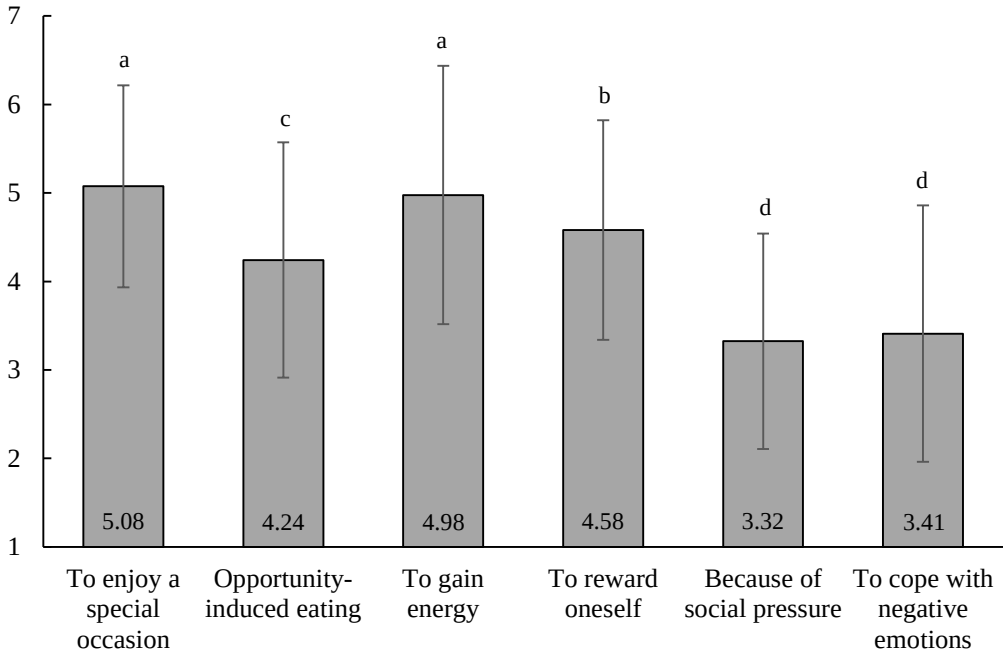


Figure 7.1. Mean acceptability of the six reasons (in order of how common they are) for unhealthy snacking provided by Verhoeven et al. (2015). Error bars represent standard deviations. Different letters indicate significant differences at the $p \leq .002$ level tested with post hoc LSD tests.

Results Study 7.2b

An ANOVA with the perceived acceptability of each of the reasons for snacking provided by Taylor et al. (2014) was conducted with the categories as a within-subjects factor. Maulchy's test indicated that the assumption of sphericity had been violated, $\chi^2(14) = 131.29$, $p < .001$, therefore the degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = 0.73$). A significant difference in perceived acceptability of the six categories of reasons was found, $F(3.66, 542.29) = 52.53$, $p < .001$, $\omega^2 = 0.20$. Means, standard deviations, and the results of the LSD post hoc comparisons are displayed in Figure 7.2.

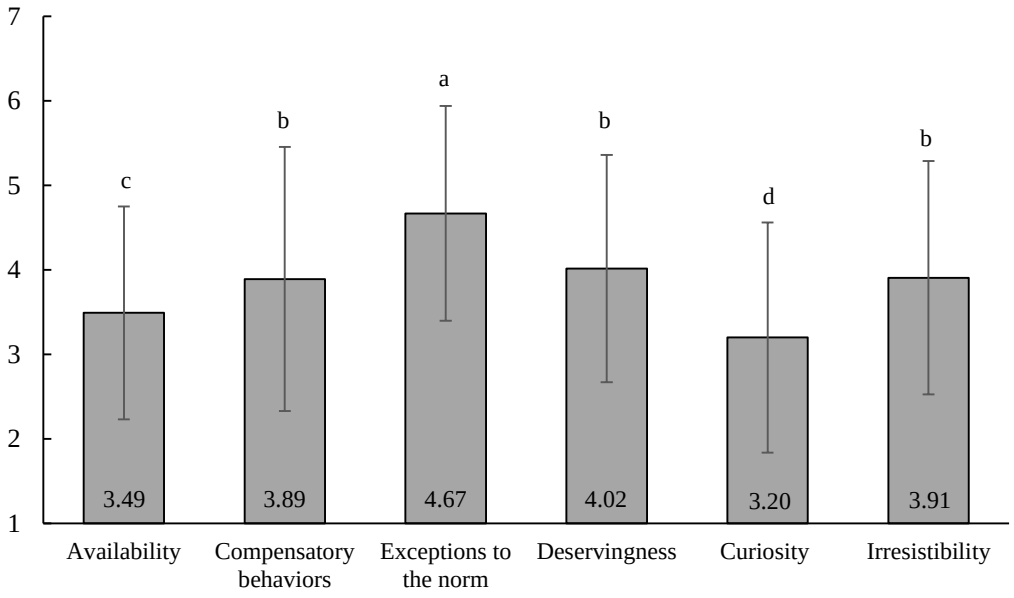


Figure 7.2. Mean acceptability of the six reasons (in order of how common they are) for unhealthy snacking provided by Taylor et al. (2014). Error bars represent standard deviations. Different letters indicate significant differences at the p <.001 level tested with post hoc LSD tests.

Discussion

Previous research identified various reasons for unhealthy snacking and documented how these differ in prevalence (Taylor et al., 2014; Verhoeven et al., 2015). Our study 7.2a and 7.2b found that these reasons also differ clearly in how good they are perceived to be to break a diet. Some are more acceptable as a reason for breaking one's diet than others. Our data reveal that the 'enjoying a special occasion' and 'gaining energy' of Verhoeven et al. were found to be the best reasons. For the reasons that were identified by Taylor et al., 'exceptions to the norm' was found to be the best reason. Together, this suggests that people find it most acceptable to break a diet for a special occasion (as a special reason by definition is an exception to the normal situation).

Importantly, this goodness or acceptability of the reasons to break a diet does not follow the pattern of how frequent reasons are given for unhealthy snacking. In the data of Verhoeven et al., both "enjoying a special occasion" and "opportunity induced eating" were the most frequently mentioned reasons for unhealthy snacking, but our data show that these two reasons differ markedly in how good people find these

reasons for doing so. People find it acceptable to break a diet for a special occasion, such as the wedding of one's best friend, but not to break a diet simply because the food is there. Verhoeven et al.'s suggestion that interventions should focus on the most common reasons for breaking a diet, is thus not the full picture as one of the categories with the most common reasons is likely to be much more harmful to future motivation than the other category of reasons.

The results of our first study (people remained more motivated to continue a diet after having broken it for a good reason than a bad one) indicated that the goodness of a reason for breaking a diet can affect the motivation to resume one's diet after the violation. Thus, the what-the-hell effect is not always as likely to occur. It is important to examine what this implies for the most frequently mentioned reasons in the literature, since it turns out that they differ in reasonability. Therefore, in Study 7.3, we explored the effects of the types of reasons that people use for unhealthy snacking on future diet motivation.

Study 7.3. Celebrating a special occasion versus opportunity-induced snacking

In Study 7.3, we aimed to demonstrate that the types of reasons for unhealthy snacking that were categorized by Verhoeven et al. (2015) differently affect future diet motivation. Therefore, half of the participants read that they ate a dessert because they were enjoying a special occasion (Special Occasion condition), whereas the other half read that they ate a dessert simply because it was lying right in front of them (Opportunity-Induced Eating condition). Note that enjoying a special occasion and opportunity-induced eating are the most common reasons given for unhealthy snacking, and thus something that Verhoeven et al. argue interventions should focus on. Since the results of Study 7.2 revealed that enjoying a special occasion is a more acceptable reason for unhealthy snacking than opportunity-induced eating, we expected that participants in the Special Occasion condition would display more positive emotions, less negative emotions, and a lower decrease in further goal motivation compared to participants in the Opportunity-Induced Eating condition.

Method

One hundred and forty-nine U.S. based participants completed our study on MTurk (99 males and 50 females, M_{age} = 33.95, *SD* = 10.74). Participants were randomly assigned to the Special Occasion condition (*n* = 75) or to the Opportunity-Induced Eating condition (*n* = 74). Participants in the Special Occasion condition were asked to imagine the following situation [the Opportunity-Induced Eating condition differs only in why they have dinner. See text in brackets]:

> *It is two months until your beach holiday. You are not happy with how your body looks like nowadays. You have gained some weight over the past few months and your clothes feel tighter. You think it is time to do something about it. You decide to follow a strict diet in order to look good at the beach during your vacation. You have been right on track since you started dieting two weeks ago. Things are going according to plan.*
>
> *Today is a very special day. Your best friend is getting married! After the ceremony, you join the wedding dinner [vs. "Today you have dinner."]*
>
> *Your favorite dessert is available. It looks mouthwatering. You look at the dessert; think about how you have been dieting for the last two weeks and how indulging in this high calorie dessert would break your diet. You decide that you can make an exception and allow yourself to indulge. You eat the dessert.*

Next, participants indicated how much happiness, satisfaction, pride, frustration, guilt, and regret they would experience in this situation (1 = *not at all*, 7 = *very much*). Subsequently, participants indicated the extent to which their decision to indulge would feel like a failure and to what extent their decision to indulge would feel like a boost to continue their diet (1 = *not at all*, 7 = *very much*). Finally, participants specified how motivated they would be to continue their strict diet (1= *not at all motivated*, 7 = *very motivated*) and how likely they would be to indulge the next day (1 = *very unlikely*, 7 = *very likely*) after eating the dessert.

Results

The results are shown in Table 7.2. First, we investigated differences in emotions between the two conditions. Participants in the Special Occasion condition displayed less guilt, regret, and frustration compared to participants in the Opportunity-Induced Eating condition. There were no differences in happiness, satisfaction and pride.

Next, we looked at motivation to stick to dieting goals. Participants in the Opportunity-Induced Eating condition indicated that their decision to indulge felt more like a failure, and were more likely to indulge the next day, compared to participants in the Special Occasion condition. There was no difference in how much they thought the indulgence was a boost to continue their diet. Participants in the Special Occasion condition indicated that they would be more motivated to continue their diet and were less likely to indulge the next day compared to participants in the Opportunity-Induced Eating condition. Together, these results suggest that of the two most common reasons for breaking a diet, one (breaking a diet for a special occasion) is less detrimental for future motivation than the other (breaking a diet simply because the food is there).

Table 7.2

Means, standard deviations, and statistics of emotions and goal persistence in Study 7.2

	Special Occasion condition *n* = 75	**Opportunity-Induced Eating condition** *n* = 74			
	M (SD)	*M (SD)*	*F*	*p*	η_{p^2}
Emotions					
Happiness	4.95 (1.46)	4.92 (1.64)	.010	.919	.000
Satisfaction	5.00 (1.56)	5.07 (1.62)	.065	.799	.000
Pride	3.45 (1.54)	3.01 (1.60)	2.84	.094	.019
Guilt	4.57 (1.86)	5.36 (1.87)	6.72	.010	.044
Regret	4.39 (1.85)	5.04 (1.77)	4.79	.030	.032
Frustration	3.85 (1.80)	4.63 (1.77)	7.02	.009	.046
Goal persistence					
Indulgence perceived as a failure	4.03 (1.83)	4.92 (1.67)	9.71	.002	.062
Indulgence perceived as a boost	3.74 (1.78)	3.57 (1.85)	0.33	.570	.002
Motivation to continue diet	5.91 (1.16)	5.27 (1.70)	7.18	.008	.047
Likeliness to indulge the next day	3.49 (2.06)	4.19 (2.15)	4.11	.045	.027

Note. All variables ranged from 1 *(not at all)* to 7 *(very much).*

Discussion

Our findings show that participants who imagined consuming an unhealthy dessert because they are enjoying a special occasion thought they would experience less negative emotions, less feelings of failure, more motivation to continue with their diet, and a lower likelihood to indulge compared to participants who imagined consuming an unhealthy dessert because the opportunity arose. Participants in the special occasion condition indicated that they would not display more positive emotions after imagining eating the dessert compared to participants in the opportunity induced eating condition. In addition, both conditions did not differ in how much they felt that eating the dessert would serve as a boost to continue with their diet.

Taken together, our findings demonstrated that the most common reasons for unhealthy snacking have different motivational consequences. This implies that health-related interventions should not solely focus on the frequency of reasons, but also on the unacceptable reasons that negatively affect long-term goal commitment.

General discussion

We started this chapter with the question of whether and when people lose their motivation after breaking their diet. We expected that the types of reasons that people have for breaking their diet would play a key role in determining whether they will be motivated to resume their diets. The results of our studies clearly show that reasonable reasons for indulgence have less negative consequences for goal motivation than unreasonable reasons for indulgence (Study 7.1) and that reasons for unhealthy snacking that have been identified in the literature differ in how reasonable they are (Study 7.2a and 7.2b). Our data also reveal that the most common reasons for unhealthy snacking (as identified by Verhoeven et al. (2015) and Taylor et al. (2014): enjoying a special occasion versus opportunity-induced eating) have different motivational consequences (Study 7.3).

The present findings point to the importance of also considering the qualitative consequences rather than solely the quantitative frequencies of the targeted behavior when developing (health-related) interventions. Similar to the legislator who has to

decide whether to focus on punishing driving violations that are most frequently displayed versus driving violations that lead to the most fatal accidents, health care professionals should carefully consider which behaviors they want to target in their interventions. Reasons that have important consequences *and* are frequent seem most important.

Coelho do Vale, Pieters, and Zeelenberg (2015) found that implementing a priori moments of goal-relaxation can increase the likelihood of long-term focal goal attainment. Thus, individuals experience less difficulties sticking to their diet when planning moments when they do not have to adhere to their strict dieting rules. In a similar vein, Liu, Haws, Lamberton, Campbell and Fitzsimons (2015) found that occasional indulgences can promote healthy eating. Based on our findings, we expect that these moments of goal-relaxation or occasional indulgences are most beneficial for further goal attainment when planned at a moment when one can provide a reasonable reason, like a enjoying a special occasion.

A potential drawback of our investigation of reasons for unhealthy snacking is that we only focused on situations where individuals provide one clear reason for giving in to temptation. Taylor et al. (2014) found that people sometimes use multiple justifications prior to indulgence (e.g., I ate the cake because it was my friend's wedding and because it was lying in front of me). However, we think that this is not problematic for our studies since people often have a predominant reason to rationalize their goal violations (e.g., the main reason for eating the cake is celebrating my friend's wedding). People having one reason, after all, are more likely to act upon that reason than people who have a disjunction of reasons (Shafir et al., 1993).

Another potential limitation is that in Study 7.3, we operationalized the categories of reasons detected by Verhoeven et al. (2015) into two specific scenarios. Note that we developed this study to demonstrate that having different reasons for unhealthy snacking can have different motivational consequences. Future research could investigate how different operationalizations of the reasons affect future motivation. In addition, future research could explore whether the reasons could be classified into

broader categories based on underlying dimensions, such as internally versus externally generated reasons.

The results indicated that the negative effects of breaking one's diet vary. In our studies, we looked at the immediate consequences of breaking a diet. Since dieting is a continuous goal, it would be very interesting to investigate the effects of providing different reasons for breaking a diet on long-term goal motivation. It may be the case that reasonable reasons not only buffer detrimental effects on one's diet, but also cause positive effects that sustain over time. Modern technologies[33], such as smartphones and tablets, enable researchers to perform ecological momentary assessments of dieters' goal violations (e.g, Hofmann, Baumeister, Förster, & Vohs, 2012). Future research could benefit from these technologies to examine the longer-term consequences of providing different reasons for indulgence in a natural setting.

Conclusion

The studies in the present chapter serve as an initial effort to explore whether the different reasons for breaking a diet play an important role in determining whether one will ultimately succeed in regaining motivation to lose weight. Our results show that some (frequent) reasons for breaking one's diet may actually be good reasons that do not result in a loss of motivation: Sometimes people simply have reasonable reasons to indulge.

[33] Online technologies have also been proven effective in the clinical domain (Blanken, Leusink, Van Diest, Gijs, & Van Lankveld, 2015).

CHAPTER 8

Discussion

In the previous 215 pages I have described my and others' research on self-licensing theory. Now it is time to take score and see what we have learned. The first part of this dissertation presented a state-of-the-art overview of published and unpublished self-licensing research. In this part, the effect size of licensing, possible theoretical moderators, and the robustness and replicability of the effect were investigated. The results were not very rewarding, neither promising. The different chapters casted serious doubt on the whole phenomenon of self-licensing. The second part of this dissertation has been much more rewarding, and presents novel theoretical perspectives on self-licensing and empirical tests of those. I examined the ways in which self-licensing can come about in daily life, provided an alternative account of self-licensing, and revealed the consequences of different (licensing-related) reasons for goal-incongruent behavior. Now, in this final chapter, I will summarize and integrate the findings from these empirical chapters and discuss them in relation to the gaps in the literature that were identified in the introductory chapter. The implications for self-licensing theory, the associated practical implications, and potential fruitful lines of future research are discussed.

The state-of-the-art of self-licensing

Chapters 2, 3, and 4 provided a state-of-the-art overview of self-licensing. In Chapter 2, a meta-analysis was conducted to establish the mean effect size of moral licensing (i.e., self-licensing in the moral domain) and to advance the existing theoretical framework by analyzing several moderators. The data for this meta-analysis were collected through an extensive literature search on moral licensing using the definition of Merritt et al. (2010, p. 344): "Past good deeds (or good intentions) liberate individuals to engage in behaviors that are immoral, unethical, or otherwise problematic, behaviors that they would otherwise avoid for fear of feeling or appearing less moral".

The meta-analysis, including 91 studies with a total of 7397 participants, showed that the size of the moral licensing effect is a Cohen's *d* of 0.31, which is a small-to-medium effect size (Cohen, 1992). The analysis could not confirm any of the moderators that were theorized to be of importance. Specifically, the moderator analyses did not show

that moral licensing conditions (e.g., recalling moral behavior) contrasted with negative control conditions (e.g., recalling immoral behavior) resulted in larger effects than licensing conditions contrasted with neutral control conditions (e.g., recalling neutral behavior, such as doing groceries). This was suprising in light of the moral cleansing effect, which implies that people who recall immoral behavior feel the need to compensate for the negative feelings through displaying subsequent moral behavior (Conway & Peetz, 2012; Jordan, Mullen, & Murnighan, 2011; Sachdeva, Iliev, & Medin, 2009; Zhong & Liljenquist, 2006). The meta-analysis did also not confirm the prediction in Conway and Peetz (2012) that recalling good actions leads to licensing effects, whereas recalling good traits results in consistency effects (i.e., good behavior follows good traits).

In addition, the meta-analysis did not find a difference between hypothetical and actual behavior, and it did not find support for the prediction that licensing effects are larger when the initial and subsequent behavior occur on the same domain (compared to in a different domain). Strikingly, the only moderator that was confirmed was that published studies had larger effects (average d = 0.43) than unpublished studies (average d = 0.11). This finding was in line with the funnel plot that indicated the presence of a positive publication bias. Furthermore, a post-hoc power analysis revealed that based on the effect size estimate, the included studies on average only have 28% predictive power.

Chapter 3 further focused on the evidential value of the moral licensing hypothesis by conducting a novel method for meta-analyses, a *p*-curve analysis (Simonsohn, Nelson, & Simmons, 2014a,b), on the published self-licensing tests that were included in the meta-analysis in Chapter 2. The results of this analysis indicated that research on moral licensing lacks evidential value; based on the distribution of *p*-values, the published moral licensing tests did not provide support for the hypothesis that previous moral behavior leads to behaviors that are immoral, unethical, or otherwise problematic.

The results from Chapter 3 may seem inconsistent with the results from the meta-analysis in Chapter 2, which show a small but significant licensing effect. A meta-analysis is a traditional approach to estimate the size of an effect and to test its

moderators, but does not correct for publication bias. *P*-curve is a relatively new method to provide an unbiased estimate of an effect, but since it has been introduced very recently the possible limitations of this method have not been identified. The meta-analysis from Chapter 2 revealed a positive publication bias, which could have led to an overestimation of the effect size. The *p*-curve analysis is not influenced by such a publication bias, which could explain the different results of both chapters. It is important to note that both methods are needed to draw a complete picture of an effect: Meta-analyses can analyze moderators, whereas *p*-curve analyses can provide an unbiased estimate of an effect.

Chapter 4 included three attempts to replicate the moral licensing effect by Sachdeva et al. (2009). The original authors found that writing about positive traits led to lower donations to charity and decreased cooperative behavior. They also found the opposite effect (i.e., moral cleansing): Writing about negative traits led to more charity donations and increased cooperative behavior. Based on these findings, it was proposed that moral-licensing and moral-cleansing effects can occur convergently as part of a moral self-regulation process. Following our own call to replicate important earlier licensing studies with larger sample sizes to verify the robustness of the moral licensing effect, three studies were conducted. Study 4.1 and 4.2 aimed to replicate their findings in larger samples (95% power based on the initial findings, $N_{\text{Study4.1}}$ = 105, $N_{\text{Study4.2}}$ = 150). Both studies did not confirm the original moral licensing effect. Study 4.3, which used an even larger sample (95% power based on the meta-analysis in Chapter 2, N = 940), did not confirm the original licensing effect, but did find a moral cleansing effect in that the recall of negative traits increased subsequent moral behavior. Thus, in none of these three studies, the original moral licensing effect was replicated. A meta-analysis including the original and the replicated effects revealed that the mean Cohen's *d* effect size was 0.07. In other words, there was no significant moral licensing effect across all studies.

Taken together, these chapters paint a rather bleak picture of the licensing effect. The small-to-medium effect size of moral licensing that was obtained in the meta analysis, the lack of confirmation of the theoretical moderators, the finding that the effect is

larger for published than for unpublished work (all in Chapter 2), the lack of evidential value for the licensing hypothesis in the *p*-curve analysis (Chapter 3) and the three failures to replicate the effect (Chapter 4) imply that the existence of the effect should be called into question. However, it is important to note that the meta-analysis, the *p*-curve analysis, and the replication studies focused on experiments that tested for the moral licensing effect in the 'traditional way'. It is typically assumed self-licensing consists of two consecutive behaviors or events, where initial good behavior leads to less desirable behavior. This reflects the way self-licensing was investigated in the studies that were included in the meta-analysis and *p*-curve analysis: Initial good behaviors were manipulated and the effects on less desirable subsequent behaviors were measured. For instance, recalls of good or moral behavior (Conway & Peetz, 2012; Jordan et al., 2011; Sachdeva et al., 2009) were used to manipulate the effect. Perhaps the way licensing was induced in these experiments does not solely reflect how the effect operates. After all, self-licensing is suggested to occur when after a prior good deed the chance of subsequent undesirable behavior is higher. In essence, it is just a description of a state of affairs, not a process.

The results in the first part of this dissertation thus suggest that there is very little evidence for self-licensing in the experiments that have been published so far. While it is tempting to conclude that self-licensing does not exist, I believe this is not the right conclusion. The second part of this dissertation provides a fresh start for the study of self-licensing. The necessary next steps were to get a better understanding of whether and how licensing occurs in daily life and to consider an alternative account of self-licensing.

Novel perspectives on self-licensing[34]

Despite the fact that these early chapters were not quite supportive of the licensing effect, I continued my efforts. These were for a large part inspired by the fact that most

[34] Note that the first part of this dissertation predominantly focused on licensing in the moral domain (i.e., moral licensing), whereas the second part mostly focused on licensing related to goal-directed self-regulatory behaviors. As outlined in Chapter 1, self-licensing can lead to a broad spectrum of undesirable behaviors, both at the individual and societal level. I believe that for both the social and the individual

people easily understand what licensing is when I discuss it with them, and that they readily recall such events from their own life. Therefore, novel ideas and new perspectives on self-licensing were needed. I described my attempts in Chapter 5 and 6. Chapter 5 attempted to investigate the different ways in which self-licensing can come about. Since previous research on self-licensing predominantly studied the phenomenon in controlled lab-experiments, and Chapters 2, 3, and 4 revealed that the existence of the self-licensing effect as it has previously been investigated should be called into question, Chapter 5 explored two different ways in which licensing can come about in daily life. Specifically, it was suggested that self-licensing can occur when a prior good deed makes one more likely to engage in subsequent questionable behavior ('good deed licensing') and when the temptation to display undesirable behavior initiates a search for a license ('temptation-based licensing'). In support of this distinction between two ways of licensing, participants in Study 5.1 categorized their own recalled episodes of licensing as either 'after displaying good behavior, I felt that I had permission to display certain behaviors' (good deed) or 'when I wanted to display certain behaviors, I justified these behaviors through my previous good behaviors' (temptation-based). Many participants indicated that their self-licensing episode resembled good deed self-licensing (53.8%) or temptation-based self-licensing (37.0%).

Study 5.2, in which participants were specifically instructed to recall one of the two ways of licensing, showed that participants could recall both good deed licensing (76.9%) and temptation-based licensing (73.9%), and that temptation-based licensing is more likely to be activated when the undesirable behavior is both more tempting and (marginally) more negative. In both studies, participants experienced more negative emotions after temptation-based licensing experiences compared to good

---

domain, the underlying processes of the licensing effect are very similar. Importantly, research on moral self-regulation (Zhong, Liljenquist, & Cain, 2009) shows that similar to other self-regulatory behaviors, moral behavior is characterized by internal regulations where the desired 'moral self' motivates goal achievement, and not living up to moral aspirations leads to emotional distress (Higgins; 1987; 1996).

deed licensing experiences. The findings from Chapter 5 support the proposition that self-licensing can be elicited in (at least) two ways and show that both ways have different antecedents and affective consequences, signifying the importance of distinguishing between the different ways in which self-licensing can be triggered.

Importantly, the findings of Study 5.1 also revealed that 95.3% of the participants could recall a licensing episode, and of those 89.3% indicated that the latest episode took place in the last month. This testifies to the importance of self-licensing theory by the frequency with which it occurs. These numbers are likely even underestimations, as these are only the examples of conscious licensing experiences, and some licensing also likely occurs without one being aware of it (as suggested by Khan and Dhar (2006)).

Chapter 6 continued this search for novel approaches to self-licensing. In this chapter I suggested that reasons play a key role, and I built on motivated reasoning (Kunda, 1990) and reason-based choice theory (Shafir, Simonson, & Tversky, 1993). These theories posit that when individuals have a preferred conclusion, they want to construct justifications that are supportive of that particular conclusion. This idea was applied to temptations: It was proposed that the temptation to display undesirable behavior creates the desire to conclude that indulgence is acceptable, which in turn influences the reasoning process. Specifically, it was hypothesized that reasons are regarded as more compelling when a temptation is stronger. Accordingly, the results of four studies showed that individuals find a large variety of reasons for giving in to temptation more acceptable when exposed to a tempting compared to a less tempting situation.

Specifically, Study 6.1 and 6.2 showed that the more people think a given behavior is tempting, the more acceptable they think a variety of reasons is to justify indulgence. Interestingly, Study 6.3 revealed that people think that their own previous good deeds are more acceptable reasons for indulgence when the temptation is stronger. This finding served as a first step toward an alternative account of licensing where the temptingness of the undesirable behavior initiates a search for a license to give in to temptation. Study 6.3 shows that this license can consist of previous good behaviors, but, as Study 6.4 shows, prior negative feelings can also serve as a license. These results

thus suggest that when people are tempted to display certain behaviors, they search for justifications and any reason may serve as one.

The findings from Chapter 6 are especially interesting in the light that previous literature on self-licensing typically interprets it as a good deed leading to a bad deed. A reverse process, where the temptation to display undesirable behavior makes previous good behavior seem as a better reason to indulge, is possibly just as likely to occur. This temptation-based reasoning model seems in line with the temptation-based way of self-licensing account that was identified in Chapter 5.

Chapter 6 showed that temptations increase the acceptability of reasons for indulgence and therefore make indulgence more likely. Of course, it varies between reasons how much justification they offer. Chapter 7 focused on such differences and tested how different reasons for indulgence affect subsequent goal-motivation. Specifically, Chapter 7 investigated the consequences of having different reasons for giving in to temptation (i.e., breaking a diet by unhealthy snacking). Study 7.1 showed that reasonable reasons for unhealthy snacking have less negative consequences for goal motivation than unreasonable reasons for unhealthy snacking. Furthermore, Study 7.2a and 7.2b revealed that reasons for unhealthy snacking that have been identified in the literature differ in how reasonable they are. Some of these reasons closely reflected reasoning that is typical for self-licensing, such as 'rewarding oneself' and 'deservingness'. Moreover, Study 7.3 revealed that the two categories of reasons for unhealthy snacking that are identified as the most common reasons (enjoying a special occasion versus opportunity-induced eating) have different motivational consequences. Together, the findings from Chapter 7 showed that reasons that people provide for indulgence differ in how reasonable people think they are and that reasonable reasons for breaking a diet seem to have less negative consequences for further goal motivation than less reasonable reasons.

Toward a more complete theory of self-licensing

In this section, I will explain what these findings mean for self-licensing theory. The meta-analysis showed that the licensing effect is small and the conclusion of the *p*-

curve analysis is even more dramatic: There is no evidential value for the licensing effect in the body of empirical research. This suggests that research on self-licensing faces a dead end. On the bright side, I do believe that there are vital new directions for self-licensing that we can start to explore.

First, I believe it is important to revise the theoretical framework of self-licensing so it becomes studied less as an effect, but more as a theory. When doing so, I think that one should 1) integrate self-licensing with theorizing on consistency, 2) include temptation as an important component of self-licensing theory, and 3) consider whether predictions and explanations from related theories also hold for self-licensing. I will elaborate on these thoughts below.

Licensing and consistency

What we really need is a theory of licensing that integrates the licensing process with theories and findings that show the importance of consistency in behavior (Abelson et al., 1986; Gawronski & Strack, 2012). Consistency theory seemingly predicts the opposite of self-licensing. Traditional consistency theories predict that people who are likely to perform a good deed are expected to be more likely to engage in future good behavior, as people want to appear consistent (both to themselves and to others). A prior good deed makes one see oneself as a good person, after which one becomes more likely to engage in future good behavior (Bem, 1972). Thus, just like licensing, consistency theories predict that past behavior is a main predictor of future behavior. A comprehensive theory on self-licensing should therefore incorporate knowledge on licensing with established theories on consistency.

Note that these basic consistency theories do not claim that people are always consistent, but that they have a drive to make things consistent. This has important implications, for example that when inconsistency arises people will become motivated to reduce those inconsistencies. Classic examples are balance theory (Heider, 1946) and cognitive dissonance theory (Festinger, 1957, 1964) that postulate that when one's behavior (i.e., not giving money to a poor person) is not in line with one's view of oneself ("I am good person who cares for those who need it"), this can be resolved

by reframing the situation so it becomes consistent ("the poor person would not need my money if he would be less lazy and look for a job"). People thus want to appear consistent, but can also reappraise situations to make them consistent. This is again why I think that studying the *effect* of licensing (bad behavior following good behavior) is less interesting than studying the *process* of licensing (why does the licensing effect occur?). Thus, more knowledge on this process will help to integrate licensing theory with well-established theories on consistency. One possible avenue to do so seems considering the role of temptations in licensing, as I explain next.

Licensing and temptation

Another aspect that is vital when revising the theoretical framework of self-licensing relates to the temptingness of the bad behavior. Temptations play a key role in the licensing process. This is clearly supported by the findings from Chapter 5 and 6. Note that the results from Chapter 5 revealed that although the undesirable behavior was more tempting for participants who experienced temptation-based self-licensing, for both types of licensing the average temptingness of the undesirable behavior was rather high (M = 5.46 for good deed licensing, and M = 5.74 for temptation-based licensing in Study 6.2). It is remarkable that the role of temptations is never mentioned in the literature on self-licensing (apart from De Witt-Huberts, Evers, and De Ridder (2014b)). After all, it makes sense that people do not need a license to wash their windows or to go to the gym, but they do need a license to eat a mouthwatering chocolate cake, to drink a lot of alcohol, or to buy an expensive new dress.

Thus, theorizing on licensing should not only focus on the prior good deed, but also on the temptingness of the undesirable behavior. The often unremarked role of the temptingness of the undesirable behavior may for a part explain the discrepancy between the findings of the first part of this dissertation, which revealed that there is no evidential value for the self-licensing effect, and Chapter 6, which showed that the large majority of people experiences both good deed and temptation-based self-licensing in their daily lives: In some cases, the undesirable behavior that was measured in the research paradigms may not have been that tempting for participants.

One could argue that some dependent variables, for instance making a choice between a mouthwatering chocolate bar and a boring apple (Fishbach & Dhar, 2005), may be more tempting than others, such as preferring a Black over a White person in a job hiring task (Effron, Cameron, & Monin, 2001; Monin & Miller, 2001). In fact, it seems odd to assume that people are tempted to be racist, and refrain from engaging in such behavior until they find a reason that justifies racism. This can either imply that our proposed account of temptation-based reasoning does not apply to all types of licensing or that studies using racism as the dependent measure might not be representative for the self-licensing effect. Given that the *p*-curve indicates that the prior body of empirical research on licensing does not contain evidential value, I think future research on self-licensing would do well to test for temptingness as a possible moderator, where licensing is expected to be stronger for more tempting behaviors.

The main conclusion from Chapter 5 and 6 that the undesirable behavior itself can play a key role in the manifestation of the self-licensing is an important complement to licensing theory: Temptations can trigger a search for acceptable reasons to transgress, and a prior good deed can be seen as an acceptable reason to do so. This is closely related to several well-established psychological theories that show that people like to find reasons for what they want to do, such as motivated reasoning (Kunda, 1990), reason-based choice (Shafir, Simonson, & Tversky, 1993) and moral reasoning (Haidt, 2001). These existing theories can thus provide vital insights in the processes behind self-licensing, as research that shows how motivated reasoning can be influenced might also influence self-licensing. An example is research on motivated reasoning that finds that people's ability to justify unethical behavior can be impeded by asking them to judge the justification that someone else would use to justify similar unethical behavior (Bersoff, 1999). If motivated reasoning plays a role in self-licensing, this would imply that judging the justifications that others use to license undesirable behavior can be a way to inoculate against licensing effects.

Because the tempting aspects of the undesirable behavior seem such an essential component of self-licensing, it is important to integrate knowledge on everyday temptations with licensing theory. Factors that play an important role in giving in to

temptation are self-control (Baumeister, Bratslavsky, Muraven, & Tice, 1998), higher priority goals (Fishbach, Friedman, & Kruglanski, 2003), situational factors such as the presence of others, and certain personality traits (Hofmann, Baumeister, Förster, & Vohs, 2012). It is likely that giving in to temptation happens most easily when people are tired. This would predict that after exerting self-control, people may more easily find any reason for indulgence even more acceptable, because they are less able to see the irrationality of these reasons.

A positive aspect of a temptation-based theory of licensing is that it seems more in line with theories on consistency than traditional theories on licensing. It seems that people who experience temptation-based licensing try to make their attitudes *consistent* with their desired behaviors (i.e., giving in to temptation) by searching for justifications that they find in their prior good behavior. Specifically, temptation-based licensing shows some close resemblances with cognitive dissonance reduction, which implies that individuals want to reduce the negative tension that they experience when they realize that their behaviors contradict their beliefs (Festinger, 1957). For example, when someone wants to lose weight, but is unable to resist the temptation of a mouthwatering hamburger, one can decide that it is probably not a large goal violation, with the justification that the hamburger is probably not that unhealthy ("I chose the regular hamburger rather than the more unhealthy cheeseburger"). The urge to reduce cognitive dissonance emerges in situations where people experience feelings of discomfort. This can be prior to, during, or after displaying certain behaviors. Chapter 6 showed that temptation-based licensing occurs when people are tempted, so when they are likely to indulge. Thus, temptation-based licensing can be regarded as a process to reduce cognitive dissonance in the situation prior to indulgence. This process of 'anticipated dissonance' was first described by Festinger in 1964. So, while licensing may look like inconsistent behavior, people's inherent preference for consistency can still play a large role in licensing theory. The fact that temptation-based licensing seems to have so much similarities with consistency theories, such as cognitive dissonance, is a good step towards more integrative framework of consistency and self-licensing.

Licensing and related effects

There is a variety of literature describing "negative spillovers", where a second action offsets the effectivity of a prior action or state. I will discuss some findings from that literature, as I think they can play role in constructing an integrative framework of self-licensing.

First, self-licensing seems to capture what economists call *the rebound effect*, which is a behavioral response that offsets the effects of a measure that aimed to reduce environmental impacts (Greening, Greene, & Difiglio, 2000). For instance, households are found to increase the frequency of clothes washing after receiving a high-efficiency washing machine, partially offsetting the energy that was saved by buying the more efficient machine (Davis, 2008). The rebound effect is a relabeling of '*The Jevons Effect*' (or the '*Jevons Paradox*', named after William Stanley Jevons, 1865), which posits that when technological advances increase the efficiency with which a resource is used, consumption rate of that particular resource subsequently grows as a consequence of rising demand. Second, *risk homeostasis theory* (Wilde, 1998) suggests that to maintain constant levels of overall risk, people take less risk when their environment becomes more risky, and more risk when their environment becomes safer. In other words, when wearing seat-belts people might drive faster because they feel safer in their car. A third theory is *moral hazard theory*, which predicts that having insurance leads people to take more risk (Arrow, 1963). All these theories and findings suggest that doing something better, more safe or more efficiently, does not have a full effect as people adjust their behavior that offsets (part of) the gain.

Fitting those theories to self-licensing, they posit that prior actions or states offer a license (or in these cases efficiency, protection, and security), that can lead to paradoxical adverse consequences. An example of how such similar findings of negative spillovers can be informative for research on self-licensing is the work by Shavell (1979), who found that moral hazard is predominantly driven by the motivation to prevent losses. This motive can also play a role in self-licensing; perhaps one of the motivations that people have to search for a license is the prevention of

losing for example money (by making donations) or time (by performing voluntary work).

Self-licensing also seems related to theories on *entitlement*, which posit that people can feel 'entitled' to take more than others because of their previous positions and behaviors (Campbell, Bonacci, Shelton, Exline, & Bushman, 2004; De Cremer & Van Dijk, 2005). Campbell et al. argue that entitlement is a component of narcissism. Narcissists likely label their behavior as more virtuous and good compared to other people. That is why I predict that narcissists would more often feel licensed to 'take a bigger piece of the cake'.

It is not only the case that licensing research can learn from existing theories, but it is also the case that these theories can learn from insights in licensing. For instance, the temptingness of the undesirable behavior may not only play a main role in the occurrence self-licensing, but also in the occurrence of moral hazard. One could argue that skiing off-piste is more tempting than losing one's luggage is. Given that the costs associated with both are covered by a travel insurance, moral hazard (with more risky behavior when one is insured) is likely for both. However, the temptation account of licensing would thus predict that moral hazard is more likely in the former case: the temptingness of off-piste skiing makes the search for a license more likely.

Practical implications

The initial chapter of this dissertation started with an example of self-licensing related to the Ice Bucket Challenge, in which people post videos of themselves dumping a bucket with ice water on their head to promote awareness of ALS disease and to encourage donations to ALS-related charities. The previous literature on self-licensing would predict that individuals who participated in this challenge engage in fewer future charitable actions. In this section, the practical implications of the findings of the current dissertation for these kinds of situations will be discussed.

It is not always the case that after displaying good or virtuous behavior, people feel licensed to display less desirable behavior. This is especially illustrated by the findings from Chapter 4. Therefore, I believe that people should not necessarily worry that after

making a donation via the Ice Bucket Challenge, individuals will engage in undesirable behaviors. However, in some cases, prior good deeds do lead to less desirable behaviors; this is apparent from the findings of Chapter 5 and 6. This is especially likely if people are tempted to display undesirable behavior (in this case, declining an additional donation request): They will search for a justification to give in to this temptation. Thus, when individuals do not want to make a second donation, they can refer to their prior good behavior to justify this. For some individuals, it will be less tempting not to make a donation (for instance because they have more money to spend), and even after participating in the Ice Bucket Challenge, there is a considerable chance that they will also donate money to other charities. For others, who have less money to spend, this would imply that they will probably donate less money to other charities. This is consistent with findings of Young, Chakroff, and Tom (2012) that doing good can lead to more good. Accordingly, it is important to consider multiple situational features when determining when people display undesirable (or less desirable) behaviors, rather than solely focusing on situations in which people just performed good behaviors.

To summarize, I believe that people should not necessarily worry that after good behavior, they will feel a strong urge to compensate for this good behavior through displaying less desirable behavior (as one would have done based on the previous literature on self-licensing). However, people should worry that individuals will display less desirable behavior when the vicious or goal-incongruent behavior is very tempting. Therefore, interventions aimed at stimulating virtuous and healthy behavior should not underestimate the power of temptation. For instance, when people are tempted to engage in unhealthy behaviors, justifying giving in through using prior good behaviors should become more difficult. Health practitioners can use descriptive norms (Cialdini, Reno, & Kallgren, 1990) to inform people that healthy actions, such as going to the gym and eating vegetables are normal rather than good (e.g., "75% of the people visits the gym regularly"). This should be done at critical moments when people are tempted to indulge. In this way, the good behavior 'loses' its licensing capacities, which makes it more difficult to justify the undesirable behavior. Another way of making justifying through previous good deeds more difficult is to implement

a system where people have to justify their choices to others. For instance, people who intend to save money can engage in a "commitment system" where they have to explain all their intended expenditures to other people.

The findings that temptations affect reasoning can also have important practical implications for marketing professionals. Previous research suggested that self-licensing might play a role in buying luxurious products (Khan & Dhar, 2006). As the purchase of luxurious products is often associated with feelings of guilt (Strahilevitz & Myers, 1998), people need to reduce these feelings through justifying their purchase. Since the findings from Chapter 6 show that externally provided or created reasons are actually regarded as better reasons to indulge when a product is more tempting, providing in-store advertisements with a possible justification could be more effective for more tempting products. Based on these findings one could expect that applying cause-related marketing to a hedonic product, such as the company making a small donation to a third-world-country when buying a luxury designer jeans, is more effective than applying cause-related marketing to a functional product, such as a vacuum cleaner.

Future directions

The results described in this dissertation have important implications for future research on self-licensing, which are outlined in this section. Since there is no evidential value for the self-licensing effect as it used to be studied, more research on self-licensing is highly warranted. Both the meta-analysis and the *p*-curve analysis revealed that "traditional" studies on self-licensing are highly underpowered. Accordingly, researchers studying self-licensing are strongly advised to increase the power of the studies. As a rule of thumb, in a study with a control and licensing condition, at least 165 people in each condition are required to have 80% power (see Chapter 2).

In addition, when studying self-licensing it is important to not solely focus on the prior good behavior, but also on the temptingness of the undesirable behavior, when developing paradigms to measure the self-licensing effect. Two studies from De Witt-

Huberts et al. (2014b) show a correlation between self-rated temptingness of indulgent behavior and the number of justifications that people indicate or generate to justify giving in to this temptation. Apart from our studies in Chapter 6 linking those ideas to self-licensing, research on self-licensing did not focus on the temptingness of the undesirable behavior. Therefore, future research should extend these novel insights through investigating in which domains temptation-based licensing occurs. It has been found in the domain of individual self-regulatory behaviors, but it has not yet been established whether this also plays a role in the emergence of racist attitudes, ecofriendly behaviors, cheating, and other behaviors that have harmful consequences at the societal level.

Moreover, researchers should critically reflect on the manner in which they intend to induce self-licensing, especially since individuals often try to engage in socially desirable responding (Paulhus, 1984). As discussed in Chapter 2, a more careful consideration of the dependent variables being used in self-licensing studies is essential. For example, quite some research on licensing includes scales that measure people's stated intention to want to help out (for instance the Willingness to Volunteer Scale, DeVoe & Pfeffer, 2007). On these types of scales, people may overstate their willingness to help other persons, perhaps partly due to social desirability concerns. If people in a self-licensing condition indicate a lower willingness to help others compared to people in a control condition, this could indeed reflect a licensing effect. However, this could also suggest that people in the licensing condition actually became consistent with their prior good self and became more honest. It is important that researchers studying licensing are aware of these possible confounds and address or prevent them in their studies.

Future research should in my view also focus more directly on the consequences of self-licensing. This is of particular interest because the results from Chapter 7 show that some justifications for undesirable behaviors have more negative affective and motivational consequences than others. So far, it is not clear whether self-licensing is a repetitive process, and whether some types of licensing are more likely to return than others.

To conclude

There is low evidential value for the licensing effect in the body of empirical research. This is a problem. It means that the published research does not warrant any firm conclusions about when the effect would occur and how strong it may be. Importantly, the novel ideas present in the latter half of this dissertation make it clear that the lack of evidential value in the published literature does not necessarily imply that the phenomenon of licensing does not exist: If one needs a license to give in to the temptation of undesirable behavior, it is likely that one will find such a license.

Samenvatting

'Zelf-licensing' houdt in dat mensen zichzelf toestaan om ongewenst gedrag (of minder gewenst gedrag) te vertonen *omdat* zij eerder goed gedrag hebben getoond. In het morele domein wordt dit effect ook wel 'morele licensing' genoemd; wanneer mensen eerst moreel gedrag vertonen zijn ze later eerder geneigd om immorele of onethische gedragingen te vertonen.

In het eerste gedeelte van dit proefschrift wordt onderzocht wat de grootte van het morele licensing effect is, hoe robuust dit effect is, en wat de belangrijkste moderatoren van dit effect zijn. In hoofdstuk 2 is een meta-analyse over het morele licensing effect uitgevoerd waarin 91 eerdere studies werden meegenomen. Deze analyse laat zien dat de Cohen's *d* effectgrootte 0.31 is, een klein tot middelgroot effect. Daarnaast laat deze meta-analyse geen empirisch bewijs zien voor het bestaan van verschillende moderatoren die er op basis van bestaande theorie wel zouden moeten zijn. De meta-analyse levert wel bewijs voor publicatie bias: de effectgroottes van gepubliceerde artikelen zijn groter dan de effecten van ongepubliceerd werk. In hoofdstuk 3 wordt direct op deze bevindingen voortgeborduurd door het uitvoeren van een *p*-curve analyse. Deze recentelijk ontwikkelde statistische techniek toetst het bestaan van een effect zonder beïnvloed te worden door publicatie bias. De resultaten van deze analyse laten zien dat er geen evidentie is voor het bestaan van het zogeheten morele licensing effect. In hoofdstuk 4 worden drie experimenten uitgevoerd om eerdere gepubliceerde zelf-licensing effecten te repliceren. De originele resultaten worden in deze experimenten niet gerepliceerd. Samengevat schetsen de eerste drie empirische hoofdstukken uit dit proefschrift geen rooskleurig beeld over de licensing theorie. Echter, zoals aangegeven in stelling 10 betekent het niet kunnen repliceren van een effect in experimentele paradigma's *niet* dat het effect er in de werkelijkheid niet is.

In het tweede gedeelte van dit proefschrift worden nieuwe theoretische perspectieven op de zelf-licensing theorie beschreven en onderzocht. Hoofdstuk 5 laat zien dat 95.3% van alle participanten in staat is om de manifestatie van het zelf-licensing effect in hun eigen gedrag te herinneren. Dus, ondanks het feit dat het moeilijk is om zelf-licensing effecten te vinden in experimentele settings blijken mensen dit effect wel degelijk in

hun dagelijks leven te ervaren. Daarnaast laat hoofdstuk 5 zien dat er twee manieren zijn waarop het zelf-licensing effect zich in het dagelijks leven kan manifesteren: 1) mensen vertonen goed gedrag en vinden dat zij het daarna verdienen om iets minder goeds te mogen doen en 2) mensen worden verleid om iets slechts te doen en praten het ingaan op deze verleiding goed door er eerder goed gedrag 'bij te halen'. Dus, zowel eerder goed gedrag als de verleiding om slecht gedrag te vertonen kunnen het zelf-licensing effect opwekken. In hoofdstuk 6 is geopperd dat de verleiding om ongewenst gedrag te vertonen het redeneringsproces van mensen beïnvloedt. Een viertal experimenten laat zien dat naarmate een verleiding sterker wordt, mensen irrelevante argumenten als meer acceptabel gaan zien om in te gaan op de verleiding. Deze resultaten komen overeen met het tweede type zelf-licensing dat in hoofdstuk 5 werd gevonden: Mensen worden eerst verleid om iets slechts te doen en trachten het ingaan op deze verleiding vervolgens met hun eerdere goede gedrag goed te praten. Voorheen focuste de zelf-licensing theorie zich volledig op de effecten van eerder goed gedrag en werd de optie dat het effect wordt opgewekt door de verleiding nagenoeg niet genoemd. Dus, deze hoofdstukken laten een omgekeerd proces van het zelf-licensing effect zien, waar goed gedrag niet altijd tot slecht gedrag hoeft te leiden, maar de verleiding van het slechte gedrag ertoe kan leiden dat mensen het goede gedrag als reden gaan inzetten. Hoofdstuk 7 gaat verder in op deze redeneringsprocessen en laat zien dat redenen voor het ingaan op een verleiding kwalitatief van elkaar kunnen verschillen. Hierdoor hebben sommige redenen wel een negatief effect op verdere doelmotivatie en andere redenen niet.

De literatuur over zelf-licensing moet sterk herzien worden omdat het huidige onderzoek naar het effect niet robuust blijkt te zijn. Daarnaast zullen zowel het voorgestelde onderliggende proces uit dit proefschrift alsmede gerelateerde theorieën (die in hoofdstuk 8 uitgebreid worden beschouwd) in de bestaande zelf-licensing theorie geïntegreerd moeten worden.

References

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum, P. H. (1968). *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk costs. *Organizational Behavior and Human Decision Processes, 35*, 124-140. doi: 10.1016/0749-5978(85)90049-4

Arnow, B., Kenardy, J., & Agras, W. S. (1992). Binge eating among the obese: A descriptive study. *Journal of Behavioral Medicine, 15*, 155-170. doi: 10.1007/bf00848323

Aronson, J. Cohen, G., L., & Nail, P. R. (1999). Self-affirmation theory: An update and appraisal. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp 127-148). Washington, DC: Americal Psychological Association.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review, 53*, 941-973. doi: 10.1016/b978-0-12-214850-7.50028-0

Atalay, A. S., & Meloy, M. G. (2011). Retail therapy: A strategic effort to improve mood. *Psychology and Marketing, 28*, 638-659. doi: 10.1002/mar.20404

Bastardi, A., & Shafir, E. (2000). Nonconsequential reasoning and its consequences. *Current Directions in Psychological Science, 9*, 216-219. doi: 10.1111/1467-8721.00098

Batson, C. D., & Thompson, E. R. (2001). Why don't moral people act morally? Motivational considerations. *Current Directions in Psychological Science, 10*, 54-57. doi: 10.1111/1467-8721.00114

Baucom, D. H., & Aiken, P. A. (1981). Effect of depressed mood in eating among obese and nonobese dieting and nondieting persons. *Journal of Personality and Social Psychology, 41*, 577-585. doi: 10.1037//0022-3514.41.3.577

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: is the active self a limited resource?. *Journal of Personality and Social Psychology, 74*, 1252-1265. doi: 10.1037/0022-3514.74.5.1252

Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1994). *Losing control: How and why people fail at self-regulation.* San Diego: Academic Press

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*, 351-355. doi: 10.1111/j.1467 8721.2007.00534.x

Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology*, *6*, 1-62. doi: 10.1016/S0065-2601(08)60024-6

Bersoff, D. M. (1999). Why good people sometimes do bad things: Motivated reasoning and unethical behavior. *Personality and Social Psychology Bulletin, 25*, 28-39. doi: 10.1177/0146167299025001003

Blanken, I., Leusink, P., Van Diest, S., Gijs, L., & Van Lankveld, J. D. M. (2015). Outcome predictors of internet-based brief sex therapy for sexual dysfunctions in heterosexual men. *Journal of Sex & Marital Therapy, 41*, 531-543. doi: 10.1080/0092623X.2014.931317

Blanken, I., Van de Ven, N., & Zeelenberg, M. (2012). [*A variety of studies on the self-licensing effect*]. Unpublished raw data.

Blanken, I., Van de Ven, N., & Zeelenberg, M. (2015a). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, *41*, 540-558. doi: 10.1177/0146167215572134

Blanken, I., Van de Ven, N., & Zeelenberg, M. (2015b). Good deeds and temptations: Two ways in which self-licensing can be triggered. *Manuscript under review.*

Blanken, I., Van de Ven, N., & Zeelenberg, M. (2015c). Temptation-based reasoning. *Manuscript under review.*

Blanken, I., Van de Ven, N., Spälti, A. K., & Zeelenberg, M. (2015). Reasonable reasons for indulgence. *Manuscript under review.*

Blanken, I., Van de Ven, N., Zeelenberg, M., & Meijers, M. H. C. (2014). Three attempts to replicate the moral licensing effect. *Social Psychology, 45*, 232-238. doi: 10.1027/1864-9335/a000189

Blanken, I., Van Wolferen, Van de Ven, N., & Zeelenberg, M. (2015). A *p*-curve analysis of moral licensing. *Manuscript under review.*

Bradley-Geist, J. C., King, E. B., Skorinko, J., Hebl, M. R., & McKenna, C. (2010). Moral credentialing by association: The importance of choice and relationship closeness. *Personality and Social Psychology Bulletin, 36*, 1564-1575. doi: 10.1177/0146167210385920

Brown, R. P., Tamborski, M., Wang, X., Barnes, C. D., Mumford, M. D., Connely, S., & Devenport, L. D. (2011). Moral credentialing and the rationalization of misconduct. *Ethics and Behavior, 21*, 1-12. doi: 10.1080/10508422.2011.537566

Cain, D. M., Loewenstein, G., & Moore, D. A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies, 34*, 1-25. doi: 10.1086/426699

Cain, D. M., Loewenstein, G., & Moore, D. A. (2010). When sunlight fails to disinfect: Understanding the perverse effects of disclosing conflicts of interest. *Journal of Consumer Research, 37*, 836-857. doi: 10.1086/656252

Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., & Bushman, B. J. (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment, 83*, 29-45. doi: 10.4135/9781412956253.n426

Cascio, J., & Plant, E. A. (2015). Prospective moral licensing: Does anticipating doing good later allow you to be bad now?. *Journal of Experimental Social Psychology, 56*, 110-116. doi: S0022103114001450

Chiou, W. B., Yang, C. C., & Wan, C. S. (2011). Ironic effects of dietary supplementation: Illusory invulnerability created by taking dietary supplements licenses health-risk behaviors. *Psychological Science, 22*, 1081-1086. doi: 10.1177/0956797611416253

Choi, B., Crandall, C. S, & La, S. (2014). Permission to be prejudiced: Legitimacy credits in the evaluation of advertisements. *Journal of Applied Social Psychology, 44*, 190-200. doi: 10.1111/jasp.12217

Chua, J. L., Touyz, S., & Hill, A. J. (2004). Negative mood-induced overeating in obese binge eaters: An experimental study. *International Journal of Obesity, 28*, 606-610. doi: 10.1038/sj.ijo.0802595

Cialdini, R. B. (1980). Full-cycle social psychology. In L. Bickman (Ed.). *Applied Social Psychology Annual* (Vol 1., pp 21-47). Beverly Hills, CA: Sage Publications.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58*, 1015-1026. doi: 10.1037/0022-3514.58.6.1015

Clot, S., Grolleau, G., & Ibanez, L. (2013a). *Anti-social behavior and self-licensing: Experimental evidence from a Ganster game*. Presented at 30. Journées de Micraoeconomie Appliquée, Nice, FRA (2012-06-06- 2013-06-07). Retrieved from: http://prodinra.inra.fr/record/264362

Clot, S., Grolleau, G., & Ibanez, L. (2013b). Self-licensing and financial rewards: Is morality for sale?. *Economics Bulletin*, *33*, 2298-2306.

Clot, S., Grolleau, G., & Ibanez, L. (2014). Smug alert! Exploring self-licensing behavior in a cheating game. *Economics Letters, 123*, 191-194. doi: 10/1016/j.econlet.2014.01.039

Cochran, W., & Tesser, A. (1996). The "what the hell effect": Some effects of goal proximity and goal framing on performance. In L. L. Martin & A. Tesser (Eds.), *Striving and Feeling: Interaction among goals, affect, and self-regulation* (pp. 99-120). Hillsdale, NJ: Erlbaum.

Coelho do Vale, R., Pieters, R., & Zeelenberg, M. (2015). The benefits of behaving badly on occasion: Successful regulation by planned hedonic deviations. *Journal of Consumer Psychology, in press.* doi: 10.1016/j.jcps.2015.05.001

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Conway, P., & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin, 38*, 907-919. doi: 10.1177/0146167212442394

Cornelissen, G., Bashshur, M. R. Rode, J., & Le Menestrel, M. (2013a). Rules or consequences: The role of ethical mindsets in moral dynamics. *Psychological Science, 24*, 482-488. doi: 10.1177/0956797612457376

Cornelissen, G., Soyer, E., & Karelaia, N. (2013b). Clicktivism or slacktivism? Impression management and moral licensing. In E - *European Advances in Consumer Research Volume 10*, eds. Gert Cornelissen, Elena, Teutskaja, and Ana Valenzuele, Duluth, MN: Association for Consumer Research, Pages: 244.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory, 33*, 67-80. doi: 10.1007/s00199-006-0153-z

Davis, L. W., 2008. Durable goods and residential demand for energy and water: Evidence from a field trial. *RAND Journal of Economics*, *39,* 530-546. doi: 10.1111/j.07416261.2008.00026.x

De Cremer, D., & Van Dijk, E. (2005). When and why leaders put themselves first: Leader behaviour in resource allocations as a function of feeling entitled. *European Journal of Social Psychology*, *35*, 553-563. doi: 10.1002/ejsp.260

De Hooge, I. E., Zeelenberg, M., & Breugelmans, S. M. (2007). Moral sentiments and cooperation: Differential influences of shame and guilt. *Cognition & Emotion, 21*, 1025-1042. doi: 10.1080/02699930600980874

DerSimonian R., & Laird, N. (1986). Meta analysis in clinical trials. *Controlled Clinical Trials, 7*, 177-188. doi: 10.1016/0197-2456(86)90046-2

DeVoe, S. E., & Pfeffer, J. (2007). Hourly payment and volunteering: The effect of organizational practices on decisions about time use. *Academy of Management Journal, 50*, 783-798. doi: 10.5465/amj.2007.26279171

De Witt-Huberts, J. C., Evers, C., & De Ridder, D. T. D. (2012). License to sin: Self-licensing as a mechanism underlying hedonic consumption. *European Journal of Social Psychology, 42*, 490-496. doi: 10.1111/j.1751-9004.2010.00263.x

De Witt-Huberts, J. C., Evers, C., & De Ridder, D. T. D. (2014a). "Because I am worth it": A theoretical framework and empirical review of a justification-based account of self-regulation failure. *Personality and Social Psychology Review, 18*, 119-138. doi: 10.1177/1088868313507533

De Witt-Huberts, J. C., Evers, C., & de Ridder, D. T. D. (2014b). Thinking before sinning: Reasoning processes in hedonic consumption. *Frontiers in Psychology*, *5*, 1268. doi: 10.3389/fpsyg.2014.01268

Dunning, D. (2007). Self-image motives and consumer behavior: How sacrosanct self-beliefs sway preferences in the marketplace. *Journal of Consumer Psychology, 17*, 237-249. doi: 10.1016/S1057-7408(07)70033-5

Effron, D. A. (2014). Making mountains of morality from molehills of virtue: Threat causes people to overestimate their moral credentials. *Personality and Social Psychology Bulletin, 40*, 972-985. doi: 10.1177/0146167214533131

Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring Whites. *Journal of Experimental Social Psychology, 45*, 590-593. doi: 10.1016/j.jesp.2009.02.001

Effron, D. A., & Conway, P. (2015) When virtue leads to villainy: Advances in research on moral self-licensing. *Current Opinion in Psychology, in press.*

Effron, D. A., & Monin, B. (2010). Letting people off the hook: When do good deeds excuse transgressions?. *Personality and Social Psychology Bulletin, 36*, 1618-1634. doi: 0.1177/0146167210385922

Effron, D. A., Monin, B., & Miller, D. T. (2012). Inventing racist roads not taken: The licensing effect of immoral counterfactual behaviors. *Journal of Personality and Social Psychology, 103*, 916-932. doi: 10.1037/a0030008

Effron, D. A., Monin, B., & Miller, D. T. (2013). The unhealthy road not taken: Licensing indulgence by exaggerating counterfactual sins. *Journal of Experimental Social Psychology, 49*, 573-578. doi: 10.1016/j.jesp.2012.08.012

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634. doi: 10.1136/bmj.315.7109.629

Eskine, K. J. (2012). Wholesome foods and wholesome morals? Organic foods reduce prosocial behavior and harshen moral judgments. *Social Psychological and Personality Science, 4*, 251-254. doi: 10.1177/1948550612447114

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi: 10.3758/bf03203630

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Festinger, L. (1964). *Conflict, decision, and dissonance*. Stanford, CA: Stanford University Press.

Fishbach, A. (2009). The function of value in self-regulation. *Journal of Consumer Psychology, 19*, 129-133. doi: 10.1016/j.jcps.2009.02.005

Fishbach, A., & Dhar, R. (2005). Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *Journal of Consumer Research, 32*, 370-377. doi: 10.1086/497548

Fishbach, A., Friedman, R. S., & Kruglanski, A. W. (2003). Leading us not into temptation: Momentary allurements elicit overriding goal activation. *Journal of Personality and Social Psychology*, *84*, 296-309. doi: 10.1037/0022-3514.84.2.296

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology, 4*, 195-220. doi: 10.1037/h0023552

Ganley, R. M. (1989). Emotion and eating in obesity – a review of the literature. *International Journal of Eating Disorders, 8*, 343-361. doi: 10.1002/1098-08x(198905)8:3<343::aid eat2260080310>3.0.co;2-c

Garg, N., & Lerner, J. S. (2013). Sadness and consumption. *Journal of Consumer Psychology, 23*, 106-113. doi: 10/1016/j.jcps.2012.05.009

Gawronski, B., & Strack, F. (2012). *Cognitive consistency: A fundamental principle in social cognition*. New York: Guilford Press

Gneezy, A., Imas, A., Brown, A., Nelson, L. D., & Norton, M. I. (2012). Paying to be nice: Consistency and costly prosocial behavior. *Management Science, 58*, 179-187. doi: 10.1287/mnsc.1110.1437

Greening, L. A., Greene, D. L., & Difiglio, C. (2000). Energy efficiency and consumption the rebound effect—a survey. *Energy Policy*, *28*, 389-401. doi: 10.1016/s03014215(00)00021-5

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834. doi: 10.1017/cbo978051181473.055

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random- effects models in meta analysis. *Psychological Methods, 3*, 486-504. doi: 10.1037//1082-989X.3.4.486

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology, 21*, 107-112. doi: 10.1080/00223980.1946.9917275

Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effect. *Journal of Experimental Social Psychology*, *19*, 323- 340. doi: 10.1016/0022-1031(83)90026-4

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*, 319-340. doi: 10.1037/0033-295X.94.3.319

Higgins, E. T. (1996). The "self digest": Self-knowledge serving self-regulatory functions. *Journal of Personality and Social Psychology, 71*, 1062-1083. doi: 10.1037/00223514.71.6.1062

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539-1558. doi: 10.1002/sim.1186

Hofmann, W., Baumeister, R. F., Förster, G., & Vohs, K. D. (2012). Everyday temptations: An experience sampling study of desire, conflict, and self-control. *Journal of Personality and Social Psychology, 102*, 1318-1335. doi: /10.1037/a0026545

Hofmann, W., Wineski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science, 345*, 1340-1343. doi: 10.1126/science.1251560

Hollenbeck, J. R., & Williams, C. R. (1987). Goal importance, self-focus, and the goal-setting process. *Journal of Applied Psychology, 72*, 204-211. doi: 10.1037//00219010.72.2.204

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*, e124. doi:10.1371/journal.pmed.0020124

Jackson, B., Cooper, M., Mintz, L., & Albino, A. (2003). Motivations to eat. Scale development and validation. *Journal of Research in Personality, 37*, 297-318. doi: 10.1037/e322352004-122

Jevons, W. S., 1865/1965. In: Flux, A. W. (Ed.), *The coal question: an inquiry concerning the progress of the nation, and the probable exhaustion of our coal mines*, 3rd edition 1905. Augustus M. Kelley: New York.

Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin, 37*, 701-713. doi: 10.1177/0146167211400208

Kacen, J. J. (1998). Retail therapy: Consumers' shopping cues for negative moods. In J. W. Alba & J. W. Hutchinson (Eds.), *Advances in consumer research* (Vol. 25, pp. 75-87). Provo, UT: Association for Consumer Research.

Khan, U., & Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research, 43*, 259-266. doi: 10.1509/jmkr.43.2.259

Kivetz, R., & Simonson, I. (2002). Earning the right to indulge: Effort as a determinant of customer preferences toward frequency program rewards. *Journal of Marketing Research, 39*, 155-170. doi: 10.1509/jmkr.39.2.155.19084

Kivetz, R., & Zheng, Y. (2006). Determinants of justification and self-control. *Journal of Experimental Psychology: General, 135*, 572-587. doi: 10.1037/e683162011-077

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480-498. doi: 10.1037/0033-2909.108.3.480

Langer, E. J., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of" placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology*, *36*, 635-642. doi: 10.1037//0022-3514.36.6.635

Lennox, R. D., & Wolfe, R. N. (1984). Revision of the self-monitoring scale. *Journal of Personality and Social Psychology, 46*, 1349-1364. doi: 10.1037/00223514.46.6.1349

Leonard, B. (2012). Unpublished raw data.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Liu, P. J., Haws, K. L., Lamberton, C., Campbell, T. H., & Fitzsimons, G. J. (2015). Vice virtue bundles. *Management Science, 61*, 204-228. doi: 10.1287/mnsc.2014.2053

Louro, M. J., Pieters, R., & Zeelenberg, M. (2007). Dynamics of multiple-goal pursuit. *Journal of Personality and Social Psychology*, *93*, 174-193. doi: /10.1037/00223514.93.2.174

Lowe, M. R., & Fisher, E. B. (1983). Emotional reactivity, emotional eating, and obesity: A naturalistic study. *Journal of Behavioral Medicine, 6*, 135-149. doi: 10.1007/bf00845377

Manktelow, K. (2012). *Thinking and reasoning: An introduction to the psychology of reason, judgment and decision making*. Hove, UK: Psychology Press.

Mann, N. H., & Kawakami, K. (2012). The long, steep path to equality. Progressing on egalitarian goals. *Journal of Experimental Psychology: General, 141*, 187-197. doi: 10.1037/a0025602

Marascuillo, L. A., Busk, P. L., & Serlin, R. C. (1988). Large sample multivariate procedures for comparing and combining effect sizes within a single study. *Journal of Experimental Education, 58*, 69-85. doi: 10.1080/00220973.1988.10806497

Mazar, N., & Zhong, C. B. (2010). Do green products make us better people?. *Psychological Science, 21*, 494-498. doi: 10.1177/0956797610363538

McKenna, R. J. (1972). Some effects of anxiety level and food cues on the eating behavior of obese and normal subjects: A comparison of the Schachterian and psychosomatic conceptions. *Journal of Personality and Social Psychology, 22*, 311-319. doi: 10.1097/00006199-197301000-00035

Meijers, M. H. C., Noordewier, M. K., Verlegh, P. W. J., & Smit, E. G. (2014). *Identity relevance moderates the licensing effect* (Chapter from doctoral dissertation). Retrieved from http://dare.uva.nl/record/1/432499

Merritt, A. C., Effron, D. A., Fein, S., Savitsky, K. K., Tuller, D. M., & Monin, B. (2012). The strategic pursuit of moral credentials. *Journal of Experimental Social Psychology, 48*, 774-777. doi: 10.1016/j.jesp.2011.12.017

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass, 4*, 344-357. doi: 10.1111/j.1751-9004.2010.00263.x

Miller, D. T., & Effron, D. A. (2010). Psychological license: When it is needed and how it functions. *Advances in Experimental Social Psychology, 43*, 115-155. doi: 10.1016/S0065-2601(10)43003-8

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81*, 33-43. doi: 10.1037//00223514.81.1.33

Mortensen, C. R., & Cialdini, R. B. (2010). Full-cycle social psychology for theory and application. *Social and Personality Psychology Compass, 4*, 53-63. doi: 10.1111/j.1751-9004.2009.00239..x

Mukhopadhyay, A., & Johar, G. V. (2009). Indulgence as self-reward for prior shopping restraint: A justification-based mechanism. *Journal of Consumer Psychology, 19*, 334-345. doi: 10.1016/j.jcps.2009.02.016

Mukhopadhyay, A., Sengupta, J., & Ramanathan, S. (2008). Recalling past temptations: An information-processing perspective on the dynamics of self-control. *Journal of Consumer Research, 35*, 586-599. doi: 10.1086/591105

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle?. *Psychological Bulletin*, *126*, 247-259. doi: 10.1037//0033-2909.126.2.247

Nelissen, R. M. A., De Vet, E., & Zeelenberg, M. (2011). Anticipated emotions and effort allocation in weight goal striving. *British Journal of Health Psychology, 16*, 201-212. doi: 10.1348/135910710x494952

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867-872. doi: 10.1016/j.jesp.2009.03.009

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609. doi: 10.1037/0022-3514.46.3.598

Plutchik, R. (1976). Emotions and attitudes related to being overweight. *Journal of Clinical Psychology, 32*, 21–24. doi: 10,1002/1097-4679(197601032:1<21::aidjclp2270320106>3.0.co;2-c

Polivy, J., & Herman, C. P. (1985). Dieting as a problem in behavioral medicine. In E. Katkin & S. Manuck (Eds.), *Advances in behavioral medicine* (pp. 1-37). New York: SAI

Reynolds, S., & Day, J. (1984). *Monte Carlo studies of effect size estimates and their approximations in meta-analysis*. Retrieved from http://catalogue.nla.gov.au/Record/5450328

Richard, F. D., Bond, C. F., Jr, & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*, 331-363. doi: 10.1037/1089-2680.7.4.331

Rick, S. I., Pereira, B., & Burson, K. A. (2014). The benefits of retail therapy: Making purchase decisions reduces residual sadness. *Journal of Consumer Psychology, 24*, 373-380. doi: 10.1016/j.jcps.2013.12.004

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99*, 400-406. doi: 10.1037/00332909.99.3.400

Rottenstreich, Y., & Hsee, C. K. (2001). Money, kisses and electric shocks: On the affective psychology of risk. *Psychological Science, 12*, 185-190. doi: 10.1111/14679280.00334

Ruderman, A. J. (1983). Obesity, anxiety, and food consumption. *Addictive Behaviors, 8*, 235-242. doi: 10.1016/0306-4603(83)90018-7

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science, 20*, 523-528. doi: 10.1111/j.1467-9280.2009.02326.x

Schlenker, B. R. (1980). *Impression management: The self-concept, social identity and interpersonal relations*. Monterey, CA: Brooks/Cole.

Schmeichel, B. J., & Vohs, K. (2009). Self-affirmation and self-control: Affirming core values counteracts ego depletion. *Journal of Personality and Social Psychology, 96*, 770-782. doi: 10.1037/a0014635

Schüler, S., Lehnhardt, N., & Huber, M. (2012). *Can priming prosocial values eliminate the moral licensing effect?*. Unpublished manuscript, Technische Universität Dresden, Germany.

Schwarz, N., Strack, F., Bless, H., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability

heuristic. *Journal of Personality and Social Psychology, 61*, 195-202. doi: 10.1037//0022-3514.61.2.195

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition, 49*, 11-36. doi: 10.1016/0010-0277(93)90034-s

Shavel, S. (1979). On moral hazard and insurance. *Quarterly Journal of Economics, 93*, 541-562. doi: 10.2307/18884469

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359-1366. doi: 10.1177/0956797611417632

Simonsohn, U., Nelson L. D., & Simmons, J. (2014a). *P*-curve: A key to the file drawer. *Journal of Experimental Psychology: General, 143*, 534-547. doi: 10.1037/a0033242.supp

Simonsohn, U., Nelson L. D., & Simmons, J. (2014b). *P*-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*, 666-681. doi: 10.1177/1745691614553988

Slochower, J., Kaplan, S. P., & Mann, L. (1981). The effects of life stress and weight on mood and eating. *Appetite, 2*, 115-125. 10.1016/s0195-6663(81)80005-0

Soman, D., & Cheema, A. (2004). When goals are counterproductive: The effects of violation of a behavioral goal on subsequent performance. *Journal of Consumer Research, 31*, 52-62. doi: 10.1086/383423

Spektor, M. (2014). Unpublished raw data.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology, 21*, 261-302. doi: 10.1016/s0065-2601(08)60229-4

Strahilevitz, M., & Myers, J. G. (1998). Donations to charity as purchase incentives: How well they work may depend on what you are trying to sell. *Journal of Consumer Research, 24*, 434-446. doi: 10.1086/209519

Susewind, M., & Hoelzl, E. (2014). A matter of perspective: Why past moral behavior can sometimes encourage and other times discourage future moral striving. *Journal of Applied Social Psychology, 44*, 201-209. doi: 10.1111/jasp.12214

Taylor, C., Webb, T. L., & Sheeran, P. (2014). 'I deserve a treat!': Justifications for indulgence undermine the translation of intentions into action. *British Journal of Social Psychology, 53,* 501-520. doi: 10.1111/bjso.12043

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science, 4*, 199-214. doi: 10.1287/mksc.1070.0330

Thomas, J. S., & Showers, C. (2012). Unpublished raw data.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology, 53*, 207-216. doi: 10.1016/S08954356(99)00161-4

Verhoeven, A. A. C., Adriaanse, M. A., de Vet, E., Fennis, B. M., & De Ridder, D. T. D. (2015). It's my party and I eat if I want to. Reasons for unhealthy snacking. *Appetite, 84*, 20-27. doi: 10.1016/j.appet.2014.09.013

Veldkamp, C. L., Nuijten, M. B., Dominguez-Alvarez, L., van Assen, M. A., & Wicherts, J. M. (2014). Statistical reporting errors and collaboration on statistical analyses in psychological science. *PLoS one, 9,* e114876. doi: 10.1371/journal.pone.0114876

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software, 36*, 1-48. Retrieved from http://www.jstatsoft.org/v36/i03/paper

Wilde, G. J. S. (1998). Risk homeostasis theory: an overview. *Injury prevention, 4*, 89-91. doi: 10.1136/ip.4.2.89.

Young, Y., Chakroff, A., & Tom, J. (2012). Doing good leads to more good: The reinforcing power of a moral self-concept. *Review of Philosophy and Psychology, 3*, 325-334. doi: 10.1007/s13164-012-0111-6

Zhang, P. (1999). Omnibus test of normality using the Q statistic. *Journal of Applied Statistics, 26*, 519-528. doi: 10.1080/02664769922395

Zhong, C. B., Ku, G., Lount, R. B., & Murnighan, J. K. (2009). Compensatory Ethics. *Journal of Business Ethics, 92*, 323-339. doi: 10.1007/s10551-009-0161-6

Zhong, C. B., & Liljenquist, K. A. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science, 313*, 1451-1452. doi: 10.1126/science.113072

Zhong, C. B., Liljenquist, K. A., & Cain, D. M. (2009). Moral self-regulation: Licensing & compensation. In D. De Cremer (Ed.), *Psychological perspectives on ethical behavior and decision making* (pp. 75-89). Charlotte, NC: Information Age

Merci.

Marcel en Niels, het is moeilijk om in woorden uit te drukken hoe dankbaar ik ben voor alle hulp en steun die ik de afgelopen jaren van jullie heb mogen ontvangen. Ik ga het toch proberen.

Marcel, je bent geniaal. Bedankt voor de inspirerende gesprekken die ik met je heb gehad, de inzichten die je me hebt gegeven, je kritische blik en je scherpe commentaren. Je bent niet alleen geniaal, je bent ook nog eens ontzettend aardig. Ik kon altijd bij je terecht met mijn sores, maar ook met mijn verhalen en (niet grappige) grappen. Dat ga ik ontzettend missen. Net als het warme en fijne departement dat je hebt weten te creëren. Dank voor al je wijze levenslessen, waarvan dit wellicht de meest belangrijke is:

Maak er wat van, maak er wat van, als je ontevreden bent nou doe daar dan wat an.

Niels, je bent geniaal. Het is een eer om je eerste PhD student te zijn. Bedankt voor alle fijne en leerzame gesprekken die we hebben gehad, de goede adviezen die je me hebt gegeven (excuus dat ik in ruil daarvoor kledingadvies heb gegeven), je scherpe blik en je kritische commentaren. Je bent niet alleen geniaal, je bent ook nog eens een ontzettend fijn persoon. Je zet je altijd voor iedereen in en je neemt overal de tijd voor. Je wist me op te beuren wanneer mijn project even niet zo lekker liep en gaf me er vertrouwen in dat het ging lukken. Ik ga de gezelligheid en je fijne (en strategische) adviezen erg missen.

Samen hebben jullie me de afgelopen vier jaren de beste opleiding gegeven die ik me had kunnen wensen. De lessen die ik van jullie heb geleerd zal ik altijd bij me houden en gebruiken. Merci beaucoup.

~

Yvette en Maartje, dankzij jullie voelde ik me meteen welkom bij het departement. Yvette, je bent een topper. Je bent – zonder dat je dat zelf misschien door hebt – het

zonnetje van het departement. Je bent enorm geïnteresseerd in iedereen en staat altijd voor iedereen klaar. Zo ook voor mij de afgelopen jaren, dank daarvoor. Maartje, er zijn maar weinig mensen die zo lief en behulpzaam zijn als jij. Het was super tof om bij jou op de kamer te starten en weer bij jou op de kamer onze dissertation countdown te eindigen. Zonder jou had dit boekje hier niet gelegen.

Flip, je bent slim en je eet veel. Je bent met stip de grappigste en leukste Belg die ik ken. Je hebt me enorm geholpen toen het niet zo lekker liep met mijn onderzoek, dank daarvoor. Het is een eer om je achter me te hebben staan tijdens de verdediging.

Arnoud, Ellen, Job en Terri, dank voor alle gezelligheid op het werk en daarbuiten. Ik heb genoten van de congressen, de roadtrips en de flauwe grappen.

Tila, ik heb je in een (te) korte tijd leren kennen als een hele fijne en lieve collega. Dank voor de fijne gesprekken.

Vriendinnetjes van JC Centelleo, Bibi, Bregje, Iduna, Ilse, Lotte, Machteld, Maite en Moos. Het was enorm fijn om het wel en wee over dit boekje met jullie te delen. Daarnaast was het ook enorm fijn om het met jullie over hele andere dingen dan dit boekje te hebben. Ik vond het heel bijzonder om dit jaar met het complete clubje ons 10-jarig lustrum te vieren. Proost op het volgende lustrum.

Mastervriendinnetjes, Bregje, Joëlle en Kim. Joëlle, jij raadde me tijdens de Master aan om een keer een afspraak met Marcel in te plannen. Daarmee is het allemaal begonnen, dus dank daarvoor. Jullie zijn fijne en lieve vriendinnen die altijd voor me klaar staan. Ik kan niet wachten tot iedereen weer in Nederland is.

Veronique, ik vind het heel bijzonder dat we na zo veel jaren nog zo goed bevriend zijn met elkaar. Je kent me door en door, wat heel fijn is. Bedankt voor al je steun en je lieve kaartjes.

Auk en Sanne, ook jullie mogen hier niet ontbreken. Bedankt voor de gezellige momenten. Ik hoop dat onze vriendschappen er nog heel lang mogen zijn.

Pap en mam, bedankt dat jullie altijd voor me klaar staan. Dankzij jullie ben ik wie en waar ik nu ben. Ik had me geen betere, leukere en lievere ouders kunnen wensen. Datzelfde geldt voor de broertjes en zusjes. Marije, Wouter, Christian, Mel, Florian en Manon, broers en zussen, dank voor jullie gezelligheid.

Antoon, Madeleine, Stijn, Pauline, Tjerk en Maartje, wat fijn dat ik me bij jullie zo welkom voel. Dank daarvoor.

Job, vriend, collega, paranimf en liefde van mijn leven, jij maakt me gelukkig. Wat een geluk dat ik jou in de afgelopen jaren heb leren kennen. Je bent er op alle mogelijke manieren voor me. Laten we voor altijd samen door blijven feesten.

Curriculum Vitae

Irene Blanken werd op 17 juni 1987 geboren in Ouderkerk. Daarna verhuisde zij naar Tilburg waar zij het Voorbereidend Wetenschappelijk Onderwijs afrondde aan het Theresialyceum. In 2005 begon zij met de Bachelor Cognitieve Psychologie aan de Universiteit van Maastricht. Nadat zij de keuzevakken in haar derde jaar afrondde aan de James Cook University in Cairns, Australië, begon zij in 2008 aan de Research Master Cognitive and Clinical Neuroscience, track Psychopathology, aan de Universiteit van Maastricht. Deze Master werd in 2010 opgevolgd door de Master Economische Psychologie aan de Universiteit van Tilburg. Na beide Masters Cum Laude te hebben afgerond begon zij in september 2011 als PhD student bij Tilburg Institute for Behavioral Economics Research (TIBER) en schreef zij dit proefschrift. Sinds september 2015 is zij werkzaam als onderzoeker bij het Nationaal Instituut voor Budgetvoorlichting (Nibud) in Utrecht.

Irene Blanken was born in Ouderkerk (the Netherlands) on June 17, 1987. Soon afterwards she moved to Tilburg, where she graduated from high school. In 2005 she started with her Bachelor in Cognitive Psychology at Maastricht University. After completing her electives at James Cook University, Cairns, Australia, she entered the Research Master program in Cognitive and Clinical Neuroscience, track Psychopathology, at Maastricht University in 2008. This Master was followed by a Master in Economic Psychology at Tilburg University in 2010. After successfully completing both Masters (Cum Laude), she started working as a PhD student for Tilburg Institute for Behavioral Economics Research (TIBER) in 2011, resulting in this dissertation. As from September 2015, she works as a researcher for the National Institute of Family Finance Information (Nibud) in Utrecht, the Netherlands.